DOCUMENT RESUME

ED 194 555 TM 800 684

AUTHOR	Spencer, Ernest
TITLE	Folio Assessment or External Examinations? An Investigation into Alternative Means of Assessing SCE Ordinary Grade English. Report to the Scottish Certificate of Education Examination Board from the Scottish Council for Research in Education.
INSTITUTION	Scottish Council for Research in Education.
SPONS AGENCY	Scottish Certificate of Education Examination Board, Edinburgh.
PUB DATE	79
NOTE	191p.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Criterion Referenced Tests: *English: *Evaluation Methods: Foreign Countries: High Schools: *National Competency Tests: Scoring: *Student Evaluation: Teacher Attitudes: Testing Problems: Test Reliability: *Writing Skills
IDENTIFIERS	*Scotland

ABSTRACT

From 1974 through 1978, three methods of assessing Scottish high school students' (O level) English achievement were studied: (1) Ordinary (O level) examinations: (2) assessment of writing skills (folio assessment): and (3) criterion referenced tests developed by the Scottish Council for Research in Education (SCRE) to measure the objectives of the Ordinary grade English course. Ten schools supplied complete data, and one school supplied partial data. General conclusions of the research favored the O-grade English examination because it is a well-tried method: it is used nationwide, ensuring comparability of standards: and teachers are satisfied with it. Problems with the O-grade examination include intermarker inconsistency and lack of fine discrimination among large numbers of average students. Although folio assessment could be more directly related to particular courses, inservice education in grading practices would be necessary, marker inconsistency would be present, and some teachers would be reluctant to take on the work involved. The criterion assessment would serve as an independent criterion with which to compare folio and O-grade evaluations. However, the study did not include an assessment of interpretation demanded by this method. Various appendices are attached, including some of the criterion tests and the teacher attitude questionnaire. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194555

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Folio Assessment
or
External Examination?

An Investigation into Alternative Means of Assessing SCE Ordinary Grade English

---

Report
to the
Scottish Certificate of Education
Examination Board

from the

Scottish Council for Research
in Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Fiddes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800684

2

£1·80

Folio Assessment
or
External Examination?

by

ERNEST SPENCER

An Investigation into Alternative Means of Assessing SCE Ordinary Grade English

---

Report
to the
Scottish Certificate of Education
Examination Board

from the

Scottish Council for Research
in Education

SCOTTISH CERTIFICATE OF EDUCATION
EXAMINATION BOARD

DALKEITH
1979

RESEARCHERS

Ernest Spencer

Denzil Stewart (until his death in November 1977)

Erica Thomas

Grateful acknowledgment is made of much help from other members of SCRE staff, especially JOHN POWELL, Assistant Director, and GRAHAM THORPE and SUE FRESHWATER of the Research Services Unit.

Steering Committee on Alternative Methods of Assessment of Ordinary Grade English

The research workers wish to acknowledge the support and helpful advice of the Steering Committee. The conclusions of the report are, however, those of the research team only.

J McPartlin, Esq	Alloa Academy – Convener	(resigned Sept 1975)
T A Sillars, Esq	Auchenharvie Academy, Stevenston – Convener	(from Sept 1975)
S B Smyth, Esq	CITE – Secretary	
T M Brown, Esq	Moray House College of Education	
T Meenagh, Esq	Adviser in English, Strathclyde Region	
H Ross, Esq	Adviser in English, Strathclyde Region	(from Sept 1975)
Prof A Rutherford	University of Aberdeen	(resigned July 1976)
Prof P H Butter	University of Glasgow	(from April 1977)
Miss G C Campbell	HMI	(resigned Sept 1975)
M Q Cramb, Esq	HMI	(from January 1976)
W Nicol, Esq	HMI	
J G Morris, Esq	HMCI	
J Powell, Esq	SCRE	
E Spencer, Esq	SCRE	
D Stewart, Esq	SCRE	(from Sept 1975 until Nov 1977)
Mrs E Thomas	SCRE	(from Oct 1975)
J N Gillam, Esq	SCEEB	
A J Dougherty, Esq	SCEEB	(resigned April 1975)
C P Kimber, Esq	SCEEB	(from Sept 1975)

CONTENTS

		Page
FOREWORD		5
SUMMARY		6
INTRODUCTION		8
CHAPTER I	FACT-FINDING: VIEWS AND PRACTICES, 1975	11
	I Survey of Teachers' Views	11
	II Survey of Methods of Assessment	14
	III Internal Assessment Elsewhere	14
	IV Summary of Fact-Finding Stage	15
CHAPTER II	THE INTERNAL ASSESSMENT SCHEME	16
	I Number and Choice of Schools	16
	II Type of Assessment	17
	III The Planned Folio Assessment Scheme	18
CHAPTER III	MEANS OF EVALUATION OF THE INTERNAL ASSESSMENT SCHEME	25
	I Overview	25
	II Statistical Criteria of Evaluation	25
	III Note on Terminology: Marks, Standardising, Scaling, Ranges and Bands	28
CHAPTER IV	FOLIO AND O-GRADE: RANKING AND PASS RATES	30
	I Ranking – Comparison of both Folio and O-grade with Criterion Test	30
	II Pass Rates – Comparison of Folio Standards with O-grade Standards	36
CHAPTER V	MODERATION	39
	I Visiting Moderator	39
	II Scaling	42
	III What sort of Reference Test?	49
CHAPTER VI	MODERATION: SUMMARY OF FINDINGS	53
CHAPTER VII	TWO FACTORS AFFECTING THE QUALITY OF ASSESSMENTS	55
	I Inter-Marker Inconsistency	55
	II The Influence of the Task and Choice of Tasks	67
CHAPTER VIII	SUMMARY OF STATISTICAL EVALUATION	74
CHAPTER IX	PRACTICAL PROBLEMS	76
	I Problems in Obtaining Accurate Data	76
	II Problems for Teachers	77
CHAPTER X	TEACHERS' VIEWS, 1977	82
	I The Questionnaire	82
	II Comments	86
	III Some Recommendations about Practicalities	91
CHAPTER XI	CONCLUSIONS	93

APPENDICES

Page

APPENDIX 1* : QUESTIONNAIRES USED IN SURVEY OF ENGLISH TEACHERS' VIEWS, 1975 99

APPENDIX 2 : CONSULTANTS WHO CONTRIBUTED TO THE MAKING OF THE CRITERION TEST 110

APPENDIX 3 : STAGES IN THE DEVELOPMENT OF A CRITERION TEST (including lists of English skills and areas of work covered) . . 111

APPENDIX 4 : CRITERION TEST PAPERS AND MARKING SCHEMES . 127

APPENDIX 5* : TRIAL MARKING 175

APPENDIX 6* : 'VISITING MODERATOR' SCHEME 175

APPENDIX 7* : TWO EXAMPLES OF SCHOOL RANGES, MODERATORS' RANGES AND O-GRADE RANGES COMPARED . . 175

APPENDIX 8 : EFFECTS OF SCALING IN TWO SCHOOLS WITH 'SATISFACTORY' INTERNAL ASSESSMENTS 176

APPENDIX 9* : DISTRIBUTIONS OF MARKS FOR CRITERION TEST MARKERS 183

APPENDIX 10 : DISTRIBUTIONS OF RAW MARKS, FOLIO AND O-GRADE 184

APPENDIX 11 : MEANS AND STANDARD DEVIATIONS (RAW MARKS) O-GRADE, FOLIO AND CRITERION TEST 185

APPENDIX 12 : MATCH OF AWARDS, FOLIO-O-GRADE, BY SCHOOL . 186

APPENDIX 13 : CROSSTABULATION OF CORRELATION COEFFICIENTS FOR ALL ASSESSMENT ELEMENTS IN THE PROJECT 187

**Note* (1) Appendix I, sections (a) and (c), and Appendices 5, 6, 7 and 9 are not included in this publication, but are available separately on application to the SCE Examination Board.

(2) The Council regrets the poor reproduction of the photographs on pages 128-134. This is unavoidable as the originals are no longer available.

FOREWORD

The research reported in this publication was carried out for the Board by the Scottish Council for Research in Education and the Board wishes to thank the Council for undertaking this work and in particular the Research Officer concerned, Mr Ernest Spencer, and his colleagues on the Council staff. The Board also wishes to thank the members of the Steering Committee which supervised the research. The project was funded jointly by the Board and the Scottish Education Department and the Board acknowledges the importance of the Department's interest in the project, both financial and otherwise.

The Board has decided that it would be inappropriate at this stage to attempt to implement the findings of the report in the context of the Ordinary grade examination in view of the discussions taking place regarding the Dunning Report and of the Board's moratorium on Ordinary grade syllabus revision. The Board however welcomes the report as a valuable contribution to educational debate while wishing to make it clear that neither the Board nor the Scottish Education Department necessarily subscribes to the views presented in it.

J. H. Walker

DIRECTOR
April 1979

SUMMARY

During the period 1974-1978 a research programme was carried out to investigate alternative means of assessing Ordinary grade English. The purpose was to investigate the validity, reliability, practicality, and general desirability of means of assessment incorporating a substantial element of school judgment. The experiment involved comparison of the results of assessment of English candidates by three methods: (i) the Ordinary grade examination, (ii) a system of 'folio assessment' and (iii) a 'criterion test' designed by the SCRE research team to give a comprehensive, valid and reliable assessment of the objectives of the Ordinary grade English course. Of the fourteen schools taking part ten provided full data for the experiment and one other provided partial data.

The following general conclusions may be drawn from the research:

The Ordinary grade English examination has many advantages. Its administrative arrangements are well-tried and effective, and its nationwide currency ensures comparability of standards for pupils from all schools; in this respect there is particular advantage in having a single common interpretation test. There is a general satisfaction with the Ordinary grade examination among both teachers and other users of its results. A further credit to the examination is the opportunity it affords to some candidates to show their real worth, which they may have lacked the motivation to reveal in school work; this benefit is, of course, offset by the disadvantage to pupils who work well in school but are 'bad examinees'.

The more unsatisfactory characteristics of the Ordinary grade examination, besides any undesirable influence on teaching, are threefold:

(a) it is bound to be of limited validity because it can sample only a small proportion of possible English work;

(b) like all assessment of English, it suffers, despite marker-standardisation, from serious inter-marker inconsistency;

(c) it is not a fine discriminator.

There are no obvious means of reducing the effects of these major disadvantages, short of the probably impracticable solution of multiple- or, at least, double-marking to improve marker-reliability.

If on the other hand assessment for certification at the end of S4 were wholly internal, the main advantages would be gains in validity. Teachers would probably see educational and professional advantages in internal assessment, and would be likely to feel more free in practice to teach a variety of courses and would plan them more carefully. In-service training through Trial Marking exercises would be needed.

Numerous problems would, however, arise. While many schools, with appropriate in-service training in assessment, would assess their pupils satisfactorily, a number would fail to do so for a variety of reasons. The setting of valid and comparable interpretation tests in different classes or in different schools presents great difficulties, which affect both school examinations and Folio assessment, especially the latter. Marker-inconsistency would be at least as prevalent as in the Ordinary grade examination, and the schools would be slightly more likely to 'bunch' the marks than Ordinary grade markers, so that they would prove no better discriminators. In addition, there would be some danger of schools' under-marking uncooperative pupils and over-marking some thought to be of high quality because they are in top classes. Moderation presents practical and theoretical problems. Above all, the difficulties for teachers of implementing an internal assessment scheme in current working conditions are considerable.

A combination of internal and external marks for these areas of work would in most cases be a better assessment than either alone, provided that both are validly assessing some aspects of English. It can be shown that the reliability and discrimination of writing and literature assessments improve when different single-marked assessments are added together, although the improvement is not so marked as when multiple-marking of the same script is carried out. Even when internal assessment was significantly less satisfactory than external in the SCRE project, the average of the two correlated satisfactorily with the Criterion measure.

A combined assessment scheme would, in principle, allow a reduction in the size of the external examination, since the writing task could be shorter (at least when certain types of task were set) and it would be possible to leave the assessment of literature wholly to the schools, if one were prepared to accept the disadvantage that it would be necessary to scale internal writing and literature against external writing and interpretation marks. Depending on success in developing them and their acceptance by English teachers, multiple-choice tests of interpretation might be employed. In any case, there may be a need for the assessment of reading in schools to give greater weight to response to literature and to kinds of reading other than passages similar to those used in the Ordinary grade examination.

The practical problems of internal assessment would still apply to a combined scheme, but would be more manageable without the need to set interpretation tests on the O-grade pattern or to prepare pupils for an external literature test. The researchers believe that the professional advantages to teachers in the acquisition of knowledge about, and practice in, assessment, along with the improvement in validity and reliability such a system would probably bring, outweigh the difficulties it would meet. Nevertheless, it would be advisable to reduce the burden of assessment carried by teachers by requiring internal assessment of about ten pieces of work, rather than the fourteen or fifteen demanded by the SCRE project.

Incidentally to the main thrust of its work, the project has brought to light in the Scottish context the same difficulties in defining achievement in the subject and in ensuring comparability of judgment among assessors as occur in all systems of assessment in English. The researchers suggest that there might be benefits to the *teaching* of English if teachers were more aware of these difficulties and if they shifted the emphasis of in-school assessment from assessment for discrimination to assessment to describe pupils' strengths and weaknesses in seeking to achieve the purposes of their work. The amount of teachers' very limited assessment time to be devoted to the two types of assessment depends ultimately on value judgments about the relative importance of each.

INTRODUCTION

Aims

The SCRE investigation of alternative means of assessing Ordinary grade English for the SCE Examination Board began in September 1974 with the following terms of reference quoted from the proposal for the research programme as finally revised in March 1974.

> **'Aims**
>
> 1. To report on the present practice in assessment by schools of candidates for Ordinary grade English.
> 2. To investigate and determine the optimum ways in which teachers can make assessments (possibly in the form of orders of merit) of pupil performance in selected aspects of English, these assessments being based on a wide range of work over a substantial portion of the school year.
> 3. To determine the best ways of scaling and/or moderating these assessments so that assessments of pupils in different schools may be comparable.
> 4. To investigate such other related matters as may appear in the course of the investigation to be relevant to the field of interest covered by the Board's remit.
> 5. On the completion of the above steps, to make recommendations to the Board on whether alternative means, utilising teachers' assessments, should be substituted for the whole or part of the present O-grade examination in English, to outline the administrative changes that would be involved. Costing of these changes might also be undertaken in cooperation with officials of the Board.
>
> **Objectives**
>
> 1. (a) To determine, by means of a survey, the methods at present used by presenting centres in the preparation of Order of Merit lists (Form Ex. 4) for candidates presented in Ordinary grade English. (The survey should determine the proportion of such lists which are based on one, on two, on three, and on four or more internal assessments of the candidate.)
>
>   (b) To determine the validity of such Order of Merit lists as measured by the SCE examination. (Should the results of 1.(a) allow, separate measures of validity should be determined for each of the categories of internal assessment procedures identified.)
> 2. (a) To note the objectives, as stated by SCEEB, of the Ordinary grade assessment of English as at present constituted, and to consider whether any further objectives have been implied in the O-grade papers set in the last few years.
>
>   (b) To list, after consultation with practising teachers of English, other objectives, if any, of the study of English in secondary schools which are not currently assessed.
> 3. To isolate appropriate objectives as defined in 2 for which the research programme will investigate the feasibility of internal assessment by schools.
> 4. To establish, *for the purposes of the investigation*, a broad assessment procedure that would be used to evaluate as comprehensively as possible all aspects of the performance of pupils studying English in Secondary 4 preparatory for presentation for O-grade. (This procedure would probably be too time-consuming to be employed other than for experimental purposes.)
> 5. To devise one or more practicable systems of internal assessment judged

to meet the objectives stated in 3 and employing where appropriate some of the techniques used in the assessment procedures described in 4.

6. To devise a practicable monitoring system for the internal assessment systems described in 5 and to report on the validity of such monitored teachers' assessments as measured by comparison with: (a) the relevant aspects of the SCE O-grade examination, and (b) the assessment system devised under 4 above.'

This statement of aims and objectives was interpreted by the research team, under the influence of their own educational experience and concerns and those of the various members of the Steering Committee, to imply that their work had two broad purposes, one with an assessment emphasis and one with a 'curriculum development' slant. The 'alternative means' of assessing O-grade English were taken to be 'in-school' assessment and the project did not directly consider possible changes in the external examination, though the Research Officer was co-opted on to a sub-committee of the SCEEB English Panel which, over two years, was developing, pre-testing and judging the value of Multiple-choice tests of interpretation as possible components of future examinations. As to its assessment emphasis, the project sought to provide information for SCEEB about possible forms of internal assessment, their practical feasibility and their comparability with two other assessments: the present O-grade examination and the 'comprehensive assessment procedure' designed specially for the project. At the same time the educational advantages and disadvantages for pupils were to be described and a report made of the impact of internal assessment on the teachers' attitude to developing their own courses and of the actual effects, if any, of the greater freedom allowed by internal assessment to engage in a variety of English activities. It was recognised that the project was concerned with the evaluation of a form of assessment already in fairly wide use elsewhere, but it was thought desirable to describe its effects in the Scottish context.

Procedures

The procedures adopted can be summarised in three stages, occurring in roughly chronological order, though there was some overlap between them, since one was being prepared while another was in progress.

1. A Fact-finding Stage

What was actually being done in schools in 1974-75 to provide the rank order of candidates sent by each school to SCEEB before the O-grade examination? What were teachers' views on the O-grade examination as an assessment instrument and as a means of evaluating their courses? How did they react to the idea of internal assessment? What aspects of English work were not covered by the external examination and which of these were teachers concerned with? Answers to these questions were found by survey methods, and an evaluation of assessment methods in current use was made by comparing school rank orders or marks with the 1975 O-grade results. This stage met objectives 1 and 2 in the proposal statement.

2. A Development Stage

Objectives 3. – 6. delineate the bulk of the work of the project, the experimental implementation of an in-school assessment system and the evaluation of it. The development stage, from Spring 1975 to Summer 1977, involved principally two undertakings:

(a) the establishment of a thorough and reliable assessment procedure to serve as an independent criterion with which to compare both school and O-grade assessments; this comprehensive assessment came to be called 'The Criterion Test';

(b) the planning and implementation of an internal assessment scheme and the means of monitoring it.

3. An Evaluation Stage

Judgments about the advantages and disadvantages of the scheme as an assessment procedure and as a spur to curriculum planning have been made on the basis of:

(a) comparisons between the three assessment measures as to ranking and standards;

(b) comments obtained at meetings with the teachers during the year and their responses to a questionnaire at the end of the experiment;

(c) the practical problems which arose for the teachers and the researchers.

CHAPTER I

FACT-FINDING: VIEWS AND PRACTICES, 1975

The development work of the project, towards establishing an independent Criterion Test and towards an internal assessment scheme, was carried out with some awareness of teachers' views on S4 English and of assessment methods in current use in 1974-75. Three surveys were made during that academic year: one, covering 25% of the schools presenting candidates for the SCE O-grade English examination, sought information about the means by which the schools drew up the Order of Merit list required on SCEEB's form Ex. 4, and was followed up by a study of the accuracy of school assessments as predictors of O-grade performance; the other two, between them seeking data from a further 50% of O-grade schools, were concerned with the aims of English teaching preparatory to O-grade, teachers' attitudes to the present examination and their reaction to the idea of internal assessment. The questionnaire method used in all three surveys was more suitable for obtaining the purely factual information about current assessment methods than for discovering the truth about teachers' views and their teaching purposes. The process of writing the questionnaires and the analysis of responses to them did, however, give the researchers a general impression of the state of SIV English teaching and assessment in 1975.

I. Surveys of Teachers' Views

The principal reason for drawing up the lists of 'objectives' and 'criteria' which formed the bulk of the two questionnaires was to map out for the planning of a Criterion Test the whole field of S4 English work. The main content of the questionnaires is therefore discussed in Appendix 3, which describes the development of the Criterion Test. Copies of the questionnaires are printed in Appendix 1. In both questionnaires the word 'objectives' means 'skills earning credit in assessment': the explanation for this interpretation of the term is given in the preliminary paragraphs of Appendix 3. Summaries are given here of those findings of the two surveys which seemed significant to the research team.

(*a*) *Free-response questionnaire*

The Principal Teacher of English and one other (randomly chosen) teacher from his department in 107 schools were invited to consider, comment on and add to a list of assessment criteria, and the teaching aims they imply, derived from analysis of SCE O-grade questions and markers' instructions. Strikes and work-to-rule by teachers at the time affected the response, so that only 115 teachers in 59 schools completed the questionnaire. Their comments cannot properly be said to have provided 'results' but a number of indications were given of some matters of concern.

1. As to curriculum, there was some evidence that a number of teachers would very much like a clear statement of what they should be teaching, especially with regard to grammar, correct usage, spelling and punctuation. There was, however, a broad cross-section of opinion and many emphasised what they saw as the restrictions placed by the examination on their freedom to design their own syllabuses. Nineteen of the 115 specifically referred to the 'limitedness' of the scope of the examination. The largest group of respondents in agreement with one another saw a need to develop and assess oral English (50/115), and another group

(17/115) included among their aims the development in pupils of a critical mind with regard to the mass-media.

These points apart, there was a general agreement that the SCRE analysis of O-grade criteria and aims fairly reflected what S4 classes were in fact working towards and that, on the whole, it had produced a statement of the appropriate sort of work for 16-year-olds.

2. As to opinions about methods of assessment, comments positively approving of some form of internal assessment (60/115) outnumbered a nevertheless fairly large group of responses which expressed positive approval of the existing examination (22/115). Strong disapproval of internal assessment, on the grounds that it would be impossible to standardise, was indicated by 5/115 teachers.

   There was some suspicion of criteria for good writing which referred (as do the O-grade marking instructions) to 'personality', 'forcefulness', and 'sincerity'; 20/115 teachers suggested that the assessment of these qualities, especially the last, is not possible in an examination. A group of about the same size (but not necessarily the same teachers) emphasised the importance of 'correctness' as a criterion of good English.

   An odd mixture of opinions emerged about the Reading section of the examination: some felt that the implied aims were too difficult, others that, in practice, the questions set are so 'open' that they do not really test the extent and quality of the pupil's reading at all. Since two examiners who had been consulted had been able to reconcile the 'open' questions with the aims and criteria stated in the questionnaire, perhaps there is a lack of communication between the SCEEB and teachers as to what exactly the 'Reading' questions are trying to test.

(*b*) *Fixed-response questionnaire*

Mainly on the basis of the responses in the free-response survey, a detailed list of 'objectives' for writing, interpretation and literature work was drawn up and sent to all English teachers in 109 schools (a 25% sample, stratified by Education authority and by type of school, viz: 6-Year Comprehensive, 4-Year Comprehensive, Selective). The teachers were to indicate whether they expected each objective to be achieved by pupils obtaining C pass in the O-grade, or by pupils obtaining A pass. Some items (106-115) were also designed to obtain information about attitudes to oral English, to internal assessment and to the present O-grade examination.

Five hundred and twenty-three teachers in 94 schools returned completed questionnaires. This was a good response from schools (86%). It was not possible to know exactly how many responses from individual teachers to expect, but, assuming an average of 9 teachers per school, the 523 probably constitute about 60% of the total sample. Within the limitations of questionnaire data, it may be taken that the trends of the responses of these teachers represent roughly the state of English teachers' thinking about O-grade work. When cross-tabulations were made, no consistent pattern of responses was apparent in schools of the same type or among Principal Teachers as opposed to others. The results for each item can be seen in Appendix 1, where the questionnaire is printed. Here the main trends of the teachers' choices are summarised.

It will be noted by anyone consulting Appendix 1 that no attempt was made to obtain teachers' views on the degree of sophistication with which each skill should be employed or each objective achieved. The definition of precise criteria of achievement is not possible, because performance partly depends on factors not susceptible to specification, such as emotional maturity and intellectual development. It is necessary to rely on the experience of examiners or moderators to discriminate between scripts with and scripts without sophistication. This limitation on the description of English criteria and objectives in the questionnaire should be borne in mind in considering the value of its results, as should the usual reservations about questionnaire data: the information obtained is about teachers' *views* of aims, etc., *when asked about them in this way*, not necessarily about their real aims or practices.

The following points among the views indicated seemed significant to the researchers:

1. There was general agreement that most of the very full list should be O-grade objectives.

   The skills most frequently regarded as being outwith the range of O-grade were:

   WRITING: argumentation (items 017; 019-020; 022)
   sophisticated story writing (037-043)
   poetry writing (045-046)
   using emotive and figurative language (048-049)
   some aspects of 'correctness', viz., use of the semi-colon and colon, of the dash and parentheses, spelling (056-062)

   READING: comment on the writers' techniques and skill (070; 072; 082-085)
   judging the validity of ideas in a text (080)
   some aspects of grammatical knowledge, viz., moods, participles, voices (102-105).

   On average (save for poetry writing) only about 20% of teachers omitted these items. There were, however, in addition, enough omissions of other items, which the majority had chosen, to justify the conclusion that there is some (though not much) disagreement among English teachers about the importance or practicality of different kinds of work for O-grade pupils.

2. Though most objectives were generally regarded as part of O-grade work, there was a tendency for many teachers to differentiate fairly sharply between 'C pupils' and 'A pupils' with respect to certain objectives (cf. items 016, 017, 019, 020; 046, 047, 049; 063; 070; *et al*). Some difficult objectives were thought by a number of teachers to be attainable by their 'A pupils'; e.g. item 042, write a story with a moral; or 072, give reasons for the judgment that one piece of writing is better than another. This clear differentiation between 'C achievements' and 'A achievements' is not justified by the actual difference in quality between scripts awarded, say, range 8 (=C) and range 5 (=A) in the examination. Awareness that there is some spread of ability among pupils taking O-grade may lead some teachers to underestimate the difficulty of some skills they expect of 'A candidates'. There may well be many teachers who regard the O-grade examination as an 'academic' one for which they also present 'non-academic' pupils. The impression that many teachers set higher standards than the examiners was, in fact, confirmed by the study of the comparability of school examination marks with O-grade marks (see II, below).

3. A majority thought the present examination is effective or, at least, useful for the purposes for which it is mainly used. A sizable minority, however, disagreed. (Items 109-112.)

4. There was a nearly equal division of opinion about the desirability of moderated internal assessment (Item 115). It was notable that 58·3% of the teachers consulted practised continuous assessment in S1-3, but only 21·6% did so in S4-6 (Items 113-114). (It should be noted that

'continuous assessment' was not defined in the questionnaire and may have been understood by some to mean 'periodic testing'.)

5. Seventy-two per cent of the teachers saw a need for *planned* oral work, but most thought that any assessment of it should be informal (106-108).

II. Survey of Methods of Assessment

The results of this survey and of the study of school examinations as predictors of O-grade performance in 1975 have been discussed in a separate report[1]. Very briefly, the findings were:

1. In 1974-75 there were probably no 'alternative' assessments in current use: all 104 schools surveyed used 'mock' O-grade examinations, while course work had very little or no influence, and, in any case, included quite a large proportion of Past Paper work.
2. In general, school 'mock' marks appeared to correlate quite well with O-grade marks – most correlations were between ·70 and ·80. The study did not, however, examine closely the particular characteristics of each school, nor those of the O-grade examination itself, which served as the criterion measure. It was assumed that correlations of ·70 or better indicated reasonably good prediction by the schools of the O-grade rank order. In fact, though, the size of a school's correlation would have been influenced, not only by the validity and reliability of the school examination, but also by the range of ability and number of pupils taking the two examinations (factors which varied from school to school) and by the quality of the O-grade as a test. There were at least suggestions of a good deal of inaccurate prediction in certain schools, and in certain classes within schools obtaining 'good' correlations, but at this stage there was no further exploration of these indications. The most obvious problem was judgment of the 'pass' standard – the C/D borderline. Most schools were more severe than the examiners in placing this line on their rank orders (a few were too lenient). However, the lack of match between school examination ranking and O-grade ranking for average pupils, even in schools obtaining correlations of ·70 upwards, was clear from the fact that it was common for about 12% – 15% of a school's candidates to have been 'misplaced' across the C/D borderline i.e., predicted to fail but actually passed, and vice versa – even after scaling of the school's marks to adjust its C/D borderline to as near as possible to the O-grade standard. Several schools 'misplaced' more than 20% of their pupils.

III. Internal Assessment Elsewhere

During the 'fact-finding' stage of the project, visits were made to examination boards and schools in England which were implementing internal assessment in English, and information was also obtained about Canadian, New Zealand and Australian experience of the change from an external to an internal system. The fruits of this investigation have been reported in detail in 'Internal Assessment or External Examination?' (*Teaching English*, January 1975). In summary, the perceived advantages of internal assessment were for teaching, course planning and pupil motivation (though there was some doubt about this last); the major problems of comparability and reliability and the costliness of moderating systems constituted the main disadvantages. There was a noticeable difference in commitment to internal assessment between the CSE boards' representatives and those of the GCE Joint Matriculation Board: the former tended to stress the advantages, the latter the disadvantages. It was, indeed, clear that proof of the superiority of one or other method of assessment was not available. Preference for internal or external methods was at least partly based on value-judgments about the purposes of assessment and about its role in the pupils' education. There may also have been some irrational failure to appreciate the faults of one's 'own' system, or the merits of the 'opposing' one.

---

[1] 'Methods of assessment used in making order of merit lists for the 1975 O-grade English examination and their accuracy compared with the examination.' (Unpublished report to SCEEB, E. Spencer, 1975.)

IV. Summary of Fact-Finding Stage

The English teaching world in which internal assessment was to be tried out was dominated, as far as S4 was concerned, by the O-grade examination. Methods of assessment in current use and much of the schools' course work were modelled directly on the O-grade papers. There was a general satisfaction with the examination, despite some misgivings about the 'Reading' questions and about the wording of some of the official criteria employed in assessing writing; some opinions hostile to the examination were expressed, however, because of its 'limiting' effect on the curriculum. Views on the desirability of internal assessment were evenly balanced – more so than support for the present examination might have indicated, because some teachers, though satisfied with the O-grade for some purposes, felt that internal assessment would offer new opportunities.

It seemed that, in practice, most teachers were happy to let the examination determine their S4 syllabus. Indeed, it could be argued that the O-grade examination does cover a very wide range of English skills, if one takes account of all possible options within the paper over a number of years and all possible criteria of merit employed by the examiners. (The achievements of no individual pupil are, however, assessed on such a broad basis.) It may have been this 'openness' of the examination which led some teachers to regret the absence of a clear, centrally imposed syllabus for S4 English work. Meanwhile, some of their colleagues, with apparent perversity, were objecting to the lack of opportunity for them to develop their own curricula, due to the restricting influence of the O-grade – in particular, oral work, close reading of literature, certain types of writing and active drama were said to be under-developed. In considering the work they were actually doing, most teachers tended towards rather idealistic expectations of the performance of 16-year-old pupils, especially those whom they predicted to obtain an A pass. Yet, again by contrast, there were suggestions that less able pupils were not being stretched enough by the reading material presented to them in preparation for O-grade.

CHAPTER II

THE INTERNAL ASSESSMENT SCHEME

I. Number and Choice of Schools

The size of the sample of schools the researchers could invite to participate in the project was limited by the amount of marking of the Criterion Test which could be completed during March-April and June employing the number of markers there was money to pay. The estimated maximum marking load was about 2,500 scripts (one script = 5 papers). The idea that a larger number of schools might be asked to assess the work of only a proportion of the S4 pupils in each was rejected, since it was thought that little information would emerge from such an exercise about the problems for schools of internal assessment in a realistic context. If whole schools were in question, the research team could then choose at most 15 for the experiment. Such a small group of schools, it was realised, might not allow legitimate generalisations to be made from results, but against that misgiving could be set the following considerations:

1. there would be gains from a whole school approach to practical problems;
2. closer contact could be kept with the teachers;
3. it was desirable to have favourably disposed teachers since they were being asked, not to replace their O-grade preparation with a different scheme, but to carry an extra burden;
4. the purely administrative difficulties of dealing with many more schools would take up research time;
5. some valid judgments about national internal assessment could be made with caution, if there were no reasons for believing that the 15 schools were in any way untypical of Scottish comprehensive schools; difficulties arising in some or all of 15 schools could be expected to occur in others, and success or failure by some would point to the possibility of success or failure elsewhere.

No attempt was therefore made to choose the schools randomly, but some 'common sense' criteria were applied to achieve a reasonable balance. Schools of varying sizes were required in different geographical areas, with city, town and rural catchment areas; and including both non-denominational and Roman Catholic schools. The Steering Committee of the project had suggested that at least one of the chosen schools would be one likely to have difficulty in fulfilling the requirements of the experiment (in the event, several schools found themselves in this condition); otherwise, schools could be selected on the grounds that fairly enthusiastic co-operation might be expected, though not necessarily 100% commitment from all members of staff. One other qualification was demanded by the 'Trial Marking' scheme proposed as part of the experiment: the selected schools had to be within fairly easy travelling distance of one another, to facilitate meetings of their representatives.

Strathclyde, Fife and the Grampian Region were chosen as areas including a wide variety of schools, and letters were sent to Headteachers describing the proposals and inviting participation. In Fife and Strathclyde the choice of schools depended mainly on the Research Officer's knowledge that the Principal Teacher or the Headteacher was likely to be interested, this knowledge, in some cases, having been passed on by the region's Adviser in English or a member of the Steering Committee of the project. Aberdeen having been excluded, since cities were to be represented by the Glasgow conurbation, the Grampian schools were chosen mainly on the basis of their proximity to one another. Replacements for the several schools which declined the offer were selected on similar grounds.

By April, 1976, 14 schools constituted the experimental groups and were designated as follows:

Fife = Area 1, five schools
Schools 11, 12, 13, 14, 15.

Grampian = Area 2, four schools
Schools 21, 22, 23, 24.

Strathclyde = Area 3, five schools
Schools 31, 32, 33, 34, 35.

A sixth school in the Strathclyde group – the one chosen as likely to meet difficulties due to staffing problems – withdrew at too short notice for a willing replacement to be found, though an attempt was made at the last minute to persuade another Grampian school to join in.

This selection of 14 schools did in fact, albeit thinly, cover most of the variables one would seek to account for in a random sample. They ranged in size from about 800 pupils to about 2,000; two were city schools; seven were in large towns, four in small towns with some rural intake, and one was a village school with a largely rural catchment area; three were Roman Catholic; and two were single sex (one boys', one girls'). It was found in due course that there were in the 14 schools quite large variations in the ability range both of the pupils presented for O-grade and of the S4 group as a whole. Finally, it was also later discovered that teacher-pupil ratios in the English departments of all the schools were similar (except for schools 31 and 33, which had special temporary staffing difficulties at the time of the experiment).

The principal types not accounted for at all in the group were independent and selective schools.

II. Type of Assessment

It might have been possible to engage the schools in any of several sorts of assessment procedure. The three most obvious possibilities were: (1) school examinations, (2) course work assessment (in which the pupils' stored work is assessed at the end of the course or, sometimes, also at one or two points during it) and (3) continuous assessment, a method involving the systematic collation of marks or grades, based either on course work or on special tests, and their aggregation into a final grade.

While it can be argued that all three of these procedures can have educational value, that continuous assessment can allow monitoring of *development*, and that it is possible to define criteria of achievement for the award of various grades, a basic concern of them all is *discrimination*, and they all rely ultimately on the professional judgment of teachers to effect this. The researchers were aware that, if internal assessment is to play a significant part in education, a shift in emphasis is probably needed – away from discriminatory assessment for the world outside school and on to that (limited) area of pupils' lives in which the professional judgment of teachers is the best available guide to pupils' achievements, namely, academic performance and social behaviour *within the school context*. Teachers are probably good at judging pupils' achievements in the pupil-role within that context, and there is an important place for assessment schemes which facilitate description of pupils' achievements, definition of skills and knowledge shown, and the pooling of several teachers' judgments about pupils' schoolwork. The SCRE publication, *Pupils in Profile* (SCRE, 1977), describes such a scheme.

Assessment of the 'Pupils in Profile' type might be complemented by a concern to encourage self-evaluation by the pupils as the basis for discussion between pupil and guidance teacher about progress, motivation, relations with teachers, careers guidance, personal development, and so on. The researchers were impressed by the development of this aspect of internal assessment at the Bosworth College, Leicestershire[1], and they felt that profiles and self-assessment ought to be among the concerns of any study of internal assessment, because it is in these types of assessment that significant development would be both valuable and possible. The fact is, however, that, as yet, the shift in emphasis away from discriminatory assessment has not occurred. Assessments made using profile or self-assessment schemes in their present state of development are not validated by anything outside the educational and, perhaps, social values of the teachers in a particular school. The necessary requirement that public discriminatory assessment be comparable across schools therefore led the researchers in the end to regard profiles and self-assessment, regretfully, as of only peripheral interest.

As it was considered that adequate information had been obtained about school examinations by the 1975 study, 'Methods of Assessment . . . and their accuracy' (Spencer, 1975), Course Work and Continuous Assessment remained. The research team had been predisposed to favour the former by the successful experience of TWYLREB, in CSE assessment. In preliminary discussions some teachers were, however, asked for their reaction to the idea of a 'continuous' scheme, marks to be recorded at monthly or two-monthly intervals. The suggestion met with no enthusiasm: it was felt that there would be too much administration and too much concern with marks instead of teaching and learning. There was, on the other hand, cautious acceptance of or welcome for a folio scheme in which O-grade assessment would be of the quality of work *in toto*, while each separate piece of work might be marked on completion for whatever purpose the teacher thought fit, e.g. for immediate diagnosis of weakness, or for encouragement. Thus was the type of internal assessment settled.

III. The Planned Folio Assessment Scheme

(a) Agreed Teaching Aims

The schools were invited to prepare pupils for the O-grade English examination by working towards the achievement of aims which the teachers recognised as implicit in the requirements of that examination. Occasional use of O-grade past papers or similar ready-made tests was admitted as necessary examination practice, but the teachers were asked not to use them as part of their teaching for the aims agreed in preliminary meetings. The list of aims agreed by the teachers at preliminary meetings was as follows:

WRITING
Prose
Pupils should be able to:

FACTUAL	– organise factual material clearly and logically, e.g. in reports, instructions, letters.
PERSUASIVE	– present opinions, arguments, evidence. – write persuasively, to win support.
PERSONAL (and DESCRIPTIVE)	– write about personal experience and feelings, describe personal interests, explore imaginative resources in response to various stimuli, describe scenes.
POETIC	– write stories.

Note This categorisation is derived from J. Britton's 'What's the Use' in *Language and Education*, published for the Open University by Routledge and Kegan Paul. Britton refers to:

(1) 'Transactional' writing: to direct, question, get things done, participate in society.
(2) 'Expressive' writing: to exchange or reveal feelings and opinions, convey attitudes, reveal personality.

[1] See *Reports and Reporting*, published by the Bosworth College, Leicestershire.

(3) 'Poetic' writing: to construct linguistic artefacts – organised language – as a means of trying to explore and master the complexities of reality.

The categories obviously merge into one another when people are actually writing.

Poetry, Drama

The writing of poetry and dramatic dialogue might also be regarded as aims of S4 O-grade work.

READING

Understanding and appreciation

(1) Pupils should read as widely as possible and should devote special attention to the study of a number of selected texts (prose, poetry and drama). This study should enable them to follow the narrative line, acquire insight into characters, respond to tone and to the attitudes and ideas presented by the writer, and recognise some of the elements in the text which contribute to the impression it makes (e.g. humour, suspense, structure, setting, images, dialogue, realism, etc.).
The emphasis should be on the pupil's own perception of the meaning and tone of the text: there is no advantage in the analysis or labelling of literary devices without some emotional and intellectual appreciation of their impact and implications.

(2) Pupils should be able to understand the content of a passage of appropriate difficulty which they have not previously studied, and also the deeper or less immediate aspects of the meaning of such a passage. They should have enough knowledge of language to be able to indicate how the writer's meaning has been conveyed. They should be able to understand and summarise the whole or a substantial part of a passage of straightforward prose.

(b) Method of Assessment

The pupils' work throughout the session was retained in a folio and assessed by the teachers in ways which varied slightly in different regions. Fife and Grampian teachers worked as departmental teams, assessing all their O-grade pupils as one group; in Strathclyde, each teacher was responsible for his own class. The Strathclyde schools agreed to make three assessments – in late October, in January and in April – while the others were to make only an end of course one in April. The research team hoped to obtain by this variation some evidence about the relative validity of single assessment as opposed to cumulative or the average of several, but the intention was frustrated by the inability of three of the five Strathclyde schools actually to carry out three assessments.

In all cases, three marks were required: for Writing, out of 30, for Literature, out of 20, for Interpretation, out of 50, these weightings being the familiar ones of the three elements in the O-grade examination.

(c) 'Rules' and Minimum Contents of Folio

The desirability of a statement of minimum requirements and rules for describing the conditions under which work was done had been established by the experience of the GCE Joint Matriculation Board with its Alternative internal scheme. The 'Rules for Folio Assessment' which the SCRE project teachers agreed to are set out in the following statement, which was circulated to the schools:

RULES FOR FOLIO ASSESSMENT

The work carried out by the pupils in the course of the year might arise from various approaches to the agreed aims – theme work, unit study work, or whatever the teacher might decide. It is proposed that all the pupil's work towards the agreed aims be kept in the folio, and that the following be the minimum requirements:

WRITING

(1) Each pupil should complete *nine* assignments of 350-450 words (or the equivalent in lengthier pieces).
Six of these should be chosen by the pupil and the teacher for assessment, the selection to contain evidence of the pupil's ability in the categories of writing listed in 'Agreed Aims for S4 English'.

(2) Each composition chosen for assessment should have the rubric of the assignment attached to it, and, if possible, any stimulus material, or a description of the stimulus material.

Each composition should be labelled –
  (a) '1st draft' or 'Improved draft'.
  (b) 'Unaided' or 'Aided': the type of aid should be briefly specified, e.g. 'Following class discussion'; or, 'After other pupils' criticism'; or, 'With some guidance from teacher', etc.
  (c) 'Classwork' or 'Homework.'

(3) There is no need to avoid choosing for assessment work improved by discussion or criticism, so long as all the work chosen can be said to represent the pupil's real ability in English. However, the selection should include at least one unaided first draft of an assignment done in class.

READING

(1) The pupils should keep a record of their reading as outlined on the specimen 'Record of Reading' (see below) and this should be kept in the folio.

(2) The folio should contain at least *six* pieces of evidence of the pupil's ability to understand and respond to several aspects of some of the texts studied.
('Interpretation' questions on sections of the texts studied come into this category.)
Four of these pieces of work should be chosen for assessment.
The folio should in addition contain some general comment indicating *the pupil's own understanding of and response to* the prose, poetry and drama he has studied.

(3) The folio should contain a minimum of *six* 'unseen' interpretations of the traditional type.
Four interpretations should be assessed, including two which required the pupil to summarise all or part of the passage.

(4) In each case, work for assessment should be labelled 'Aided' or 'Unaided', 'Homework' or 'Classwork', as for 'Writing' above, and at least one of the 'unseen' interpretations chosen for assessment should be work done unaided in class.

SPECIMEN RECORD OF READING (O-grade English course, August, 1976 – April, 1977)
Name:
School:
Class:

(1) *Texts studied as part of school work.* (Titles and authors)
Prose
Poetry
Drama

(2) *Texts read in addition to the above.* (Titles and authors)
(You may include here any story, book or poem you have read, any play you have read or seen performed. If you are doubtful about the suitability of a text you want to include ask for your teacher's advice.)

ORAL

(N.B. This part of the scheme is optional, and will not be included in the comparison between the school assessments and the two external examinations.)

Teachers will be asked in January and in April 1977 to give an 'impression grading' of pupils' oral ability, based on general criteria to be supplied by the Research team. If possible, teachers other than English teachers may be asked to provide an oral assessment grade too, so that the reliability of this grading might be increased.

(d) Support for Teachers

Two means of helping teachers with assessment problems were adopted: Trial Marking and provision of a Guidance booklet.

Trial Marking was modelled on the system developed by The West Yorkshire and Lindsey Regional Examining Board (TWYLREB) for the CSE. Four meetings of school representatives were planned for June and September 1976, and January and March 1977, for the purpose of establishing common standards on scripts selected by the researchers and previously sent to each school for the department, or at least three members of it, to mark. The representative thus was able to bring to the meeting his school's assessment, and not merely his own. Although only three Trial Marking meetings were actually held, this part of the work of the project was very successful: it is reported in detail in Appendix 5.

The Guidance booklet was a compilation of advice about ways of improving the validity and reliability of assessment in English, suggested criteria for marking certain writing assignments, and exemplars of tests of response to short stories in which the questions had been derived from suggested 'purposes of reading', following a model created by the Reading Development Course team of the Open University. The booklet was late in preparation and too bulky, so that any effect on teachers' assessment during the project was probably minimal. It is hoped that its recommendations may be helpful to the teachers in the future[1].

(e) Moderation

The feasibility and effectiveness of two methods of moderation were tested.

1. Visiting Moderator

   Moderators were appointed to visit the schools on two or three occasions to monitor the internal assessment scheme. Their function was (i) to discuss with teachers difficulties arising, in particular those concerning the comparability of levels of difficulty of work set in different schools (a matter of trouble to the JMB internal assessment scheme), (ii) to moderate the standard of the school assessments by marking the folios of a sample of 20 pupils.

   The statistical tests of comparability between school and moderator were those of 'Range estimates' developed for CSE moderation, described in 'Schools Council Examination Bulletin No. 5: The Certificate of Secondary Education: School-based examinations', and used by TWYLREB. The application of these tests in the SCRE project is described in the discussion of moderation in Appendix 6.

   Two types of moderator were employed. It was assumed that in a real internal assessment system moderators would be experienced and successful O-grade markers (while examiners would probably be senior moderators). Accordingly, it was decided to appoint four experienced markers to Fife and Strathclyde, each to deal with two or three schools. In the Grampian region a teacher from each school, who was involved in the scheme and had taken full part in the Trial Marking exercises, carried out moderation of marking in another school in the group. The teachers in this group did not wish to take on the task of judging whether work set by colleagues in other schools was of the appropriate level of difficulty, so one of the researchers undertook that part of the moderators' duties.

2. Statistical Standardising

   The marks given by the teachers were scaled against those obtained by the same groups of pupils in both the O-grade and the Criterion Test.

(f) Modifications to the Planned Scheme

The experimental scheme as agreed with the participating schools in early Summer, 1976, did not proceed without modification. One Fife and one Strathclyde school withdrew in Autumn, 1976, from the commitment to provide a folio assessment. Another Strathclyde school maintained its intention to provide course-work marks until March, 1977, but was not in fact able to do so; and a third member of the Strathclyde group fulfilled the commitment only partially, sending marks for two classes out of six. (The remaining two Strathclyde schools co-operated fully and did each provide three assessments in the course of the year, as requested; no attempt was made, however, to investigate the relative comparability with the Criterion Test of these three assessments separately and combined, as it was felt that too many of the pupils for whose results this study had been planned had been lost through their schools' inability to provide marks.)

---

[1] The Guidance booklet is under revision for possible wider circulation.

In the schools which fulfilled all the demands of the project, fears arose early in the session that the agreed minimum requirement for the folios would prove more than could be met by a significant proportion of pupils. Several teachers who were trying to conform with the suggestion that past papers might not be used as teaching material became anxious about the difficulty of finding or inventing alternative interpretation material in the very limited free time they had; they were also worried about the need for pupils to have examination practice. In the event, almost all the pupils involved probably did produce the minimum amount of work over the year. (There is some doubt, because in some schools pupils were allowed to remove material from their folders to revise for the O-grade examination before the moderator scrutinised the school's assessment.) The teachers' concern in the early stages was probably due to insecurity in dealing with an unfamiliar system and to the feeling that they simply did not have enough time to teach for O-grade and try to engage in some curriculum development as well. In October, 1976, the research team sent to the schools the booklet of guidance on internal assessment, the contents of which may have caused more confusion than enlightenment: it contained suggestions for diagnostic assessment and for types of English work not normally undertaken in preparation for O-grade, as well as general advice regarding the validity and reliability of assessments. Though this booklet was meant to be a set of ideas which teachers might try out or not, as they wished, it was misunderstood by some to be a very late statement of more 'rules' which they felt they had not had time to consider and plan for. As a result of reports from the schools referring to this booklet and to the fears described above, the researchers circulated in late October, 1976, an explanatory leaflet entitled 'Some Clarifications', to reassure the teachers.

'Some Clarifications' defined the essential elements of the experiment as (i) assessment of a folder of course work arising from the school's teaching towards O-grade aims, (ii) the comparison of marks awarded in that assessment with those awarded in our Criterion Test and in the O-grade examination. Reassurance was given that the agreed minimum requirements, the request to avoid past papers and the suggestions in the Guidance booklet were all subject to the teachers' judgment about what was best for their pupils and what it was possible for the pupils to do. Adjustment of the 'rules' was permitted to this extent and it was pointed out that one of the purposes of the project was precisely to report on the amount of work it was reasonable to ask of pupils and teachers in the course of a year. The document also repeated the assurance that moderation would not in any sense be an inspection of the quality of teachers' work.

There were three further alterations to the originally projected scheme, besides those allowed by 'Some Clarifications'. Since it became obvious that everybody was finding difficulty in providing the basic necessary data, two suggestions written into the plan were quietly, though regrettably, dropped: these were the invitation to assess the pupils' oral work and the request for a record of the pupils' reading during the year. The third omission was of the 'Trial Marking' exercise scheduled for January, 1977.

Detailed accounts of practical problems which occurred and of teachers' reactions to the project can be found in Chapters IX and X.

(g) Summary of Experimental Programme

Table 2.1 sets out the scheme in full, indicating the principal modifications described in **(f)** above. The time-scales for the preparation and implementation of the programme are shown in Tables 2.2 and 2.3.

TABLE 2.1: FOLIO ASSESSMENT SCHEMES (SUMMARY)

	No. of assess-ments	*Form*	*Made by*	*Trial Marking*	*Moderation*	*Validation*
Fife – originally 5 schools; 4 provided data	1	Marks for Writing /30 Literature /20 Interpretation /50	Depart-mental team	4 planned, but only 3 carried out	(1) External moderator: A. difficulty level of work B. standards of marking. (2) Scaling against O-grade	Compare marks with those given in O-grade and in Criterion Test. Obtain teachers' views on educational advantages and disadvantages
Grampian – 4 schools, all providing data	1	As Fife	As Fife	As Fife	(1) Difficulty level of work – by researcher. Standards of marking – by teacher from within the group. (2) As Fife	As Fife
Strathclyde – 5 schools; 2 provided full data, 1 partial data	3 planned, but only 1 actually used	As Fife	Each teacher own class	As Fife	As Fife	As Fife

TABLE 2.2: TIME-SCALE FOR PREPARATION OF FOLIO ASSESSMENT

February 1976	*March/April* 1976	*June* 1976	*August/September* 1976
1st contact with school	(1) Preliminary visit to school. (2) Teachers consider written proposals. (3) Second visit to receive objections, etc. (4) Meeting of school representatives to finalise plans.	1st Trial Marking	Implementation

TABLE 2.3: TIME-SCALE FOR IMPLEMENTATION OF FOLIO ASSESSMENT

September 1976	*October* 1976	*November* 1976	*December* 1976	*January* 1977	*February* 1977	*March* 1977	*April/May* 1977
Schools get organised. 1 school provides Trial Marking scripts.	(1) 2nd Trial Marking. (2) 1st Assessment (Strathclyde).	1st Moderator's visit, to discuss levels of difficulty of work.		(1) 3rd Trial Marking (abandoned). (2) 2nd Assessment (Strathclyde).		(1) Criterion Test. (2) 4th Trial Marking	(1) Final Assessment (2) O-grade exam. (3) 2nd Moderator's visit

CHAPTER III

MEANS OF EVALUATION OF THE INTERNAL ASSESSMENT SCHEME

I. Overview

The Folio Assessment scheme was evaluated statistically as regards ranking and as regards judgment of O-grade standards at each band boundary by comparing the marks awarded by the Folio Assessment and in the O-grade examination with those awarded to the same pupils in the SCRE 'Criterion Test', an extensive, thoroughly prepared and reliably marked independent assessment. This statistical evaluation is reported in Chapter IV.

Chapter V deals with the methods of moderation tried out in the project, considering their effectiveness for imposing O-grade standards on internal assessment and also their effects on awards to individual pupils.

In the course of comparing the three sets of marks it was realised that certain related matters were of considerable importance in trying to decide whether Folio Assessment is as good as or better than the external examination. Chapter VI discusses some of these matters – in particular, marker unreliability and the problems caused for comparability of assessments by allowing variation in the tasks carried out by the pupils – and the implications for assessing Writing and Literature, on the one hand, and Interpretation, on the other, are pointed out.

In addition to purely statistical concerns the researchers were interested in the practical feasibility of an internal assessment scheme and in the educational advantages and/or disadvantages. The difficulties met in the implementation of the scheme are described in Chapter IX, and suggestions are made as to means or reducing them. Practical problems also figure prominently in Chapter X, which is devoted to the views of the teachers who took part in the experiment, but another important function of this chapter is to present the teachers' opinions about the educational value of the scheme.

It was not possible in the project to make any kind of direct assessment of the amount of progress pupils made as compared with what would have been the case in a normal year. It also proved impossible, despite the good intentions of the researchers, to survey the views of the pupils themselves about the project: very few were available to be questioned in the third term, after the O-grade examination.

II. Statistical Criteria of Evaluation

Folio Assessment could be statistically evaluated as regards *ranking* and as regards *judgment of O-grade standards* at each band boundary.

(*a*) *Ranking: the function of the Criterion Test*

It may be helpful to define two well-known but still important characteristics of good assessment. (1) It should be valid – i.e., it should actually test what it claims to test: the extent and quality of pupils' attainments in English. (2) It should be reliable – i.e., it should be as free as possible from chance factors which might influence the marks awarded.

In the project's study of internal examinations in 1975[1], the school assessment was compared only with the O-grade results. In order to compare the relative qualities of Folio Assessment and the O-grade examination, an independent criterion was required, which had to be a more thorough and reliable assessment than a normal examination, because of the known imperfections of tests like the O-grade English examination. For practical reasons, such public examinations suffer a loss of validity, because they can sample only a small proportion of a candidate's attainments. They may also, despite careful preparation and moderation, occasionally present candidates with, for instance, questions which are invalid because of ambiguity of wording which was not detected until large numbers of pupils actually tried to answer them. Or, from time to time, a passage for interpretation might be chosen which requires specialised knowledge from anyone required to understand it fully. Pre-testing of the papers would bring such faults to light, but is impractical for annual O-grade examinations.

---

[1] **Reported to SCEEB in Spencer (1975).**

The principal means of reducing the unreliability which remains in an English examination even after markers' meetings – multiple-marking to reduce inter-marker inconsistency – is also normally impractical, so that a degree of unreliability is inevitably present in an O-grade examination.

The same imperfections influence internal assessment such as the Folio Scheme employed in the SCRE project, though their relative significance is different. In folio assessment there is the possibility of greater validity than in an external examination, because more work can be assessed, but there are likely to be more difficulties with reliability and with comparability of work and standards of assessment in different schools.

The function of the Criterion Test was to provide, for the purposes of the experiment, a measure of pupils' attainments in English which would combine the extensiveness of Folio assessment with the comparability of an external examination and which could be more thoroughly prepared and marked than a teacher's assessment or even an O-grade examination. Accordingly, the requirement of the remit of the project was that the researchers should create a comprehensive test of all the 'objectives' of English teaching in S4 preparatory to O-grade, optimise its validity and reliability, and use the marks given in it as the 'criterion' against which both the internal assessments and the O-grade results might be compared. The test was to be a 'criterion' in this sense only: it was not a 'criterion referenced test'.

The development of the Criterion Test posed some theoretical and practical problems which, since they have their own interest, have been kept for a separate section of the report: the detailed account of the making of the test can be found in Appendix 3. Pages 122-124 contain a statement of the range of skills and kinds of English work which the test covered and also of those which were excluded for various reasons. The test papers themselves and the marking schemes can be found in Appendix 4. It is enough to note here that there were five papers, covering: free composition with picture or 'imaginative' verbal stimuli; two writing exercises with specific purposes and for stated audiences; two passages for 'traditional' interpretation, and three others with multiple-choice questions; a test of close reading of a complete short story, with guidance as to what to look for and (within reason) unlimited time to answer; and three questions requiring 'general' responses to, respectively, a poem, a play and a prose work chosen by the pupils themselves.

'Validity' in respect of this test meant only 'face-validity', but as high as possible. The aim was to produce a test on which two of the researchers, both experienced teachers, had expended much thought, which had been pre-tested where appropriate, and which had passed through critical evaluations of its suitability for testing the English of 16-year-olds by O-grade examiners and a number of other teachers. Throughout the consideration of validity and of English 'aims', 'objectives', 'criteria', 'skills' or 'purposes', the project relied for justification of the rightness of descriptions, definitions, test questions, and passages, on the experience as teachers and examiners of the researchers and of a fairly large group of others involved in teaching and examining whose views were sought from time to time. (The names of those contributing to the project in this way are listed in Appendix 2.) The validity of the Criterion Test as an instrument for assessing as truly as possible the achievements in English of S4 pupils was therefore underwritten by several knowledgeable English teachers and examiners who gave their approval to, or suggested modifications in, its various elements. The test could be said at least to match the O-grade in respect of the care and experience employed in its making and to improve upon the SCEEB examination in its extensiveness.

The influence on the results of the Criterion Test of chance factors was reduced as compared with O-grade or Folio assessment, because, besides the normal markers' meetings, special measures were taken in marking it: (1) all writing and literature tasks (i.e., Papers I, III, and V) were double-marked by experienced O-grade markers; further, the obligation on the pupils to write three compositions, two of which were for defined purposes and audiences, and to respond to specific questions on a set text, as well as to 'open' literature questions, was likely to increase the reliability of marking, as against an absolutely free choice of composition and of literature texts and questions; (2) marker-inconsistency was removed altogether from Paper IV, the multiple-choice test, which had been subject to vetting by a team of question setters and had been satisfactorily pre-tested; (3) the procedures adopted to standardise markers' assessments in Papers I, II, III and V – adjustment of both the mean mark awarded by each marker and the spread of his marks – could correct more unreliability than the SCEEB procedure, which adjusts the mean only. (As will be seen in the discussion of the correlation coefficients, there was however one school, 13, in which the Criterion Test was not set in the standard way, with the result that its reliability for that school is more questionable.)[1]

Though it would not always be safe to assume that Folio assessment and an external examination are measuring the same things, the school assessment, the O-grade and the Criterion Test in this case did have common teaching aims and criteria of achievement. The similarity of teaching aims is evident if the Folio assessment aims (stated on page 18) are compared with the tasks set in the Criterion Test (see page 122), both having been derived ultimately from analysis of O-grade questions and markers' instructions. Marking criteria and processes for the Criterion Test were modelled on the O-grade ones and O-grade markers were used; the Trial Markings, in which the teachers involved practised assessment, also applied O-grade criteria to the sample scripts. On the basis of the considerations of validity and reliability described above, the rank order produced by the Criterion Test could be taken to be the best of the three available. A positive correlation between either of the other two sets of marks and the Criterion Test marks may be interpreted as indicating the degree to which that test and the Criterion Test are reliably measuring the same things. If the correlation between Folio and Criterion Test is significantly better than that between O-grade and Criterion Test, the Folio assessment can be said to have reliably assessed more of whatever is measured by the Criterion Test than the O-grade did. The relative merits of Folio and O-grade as regards ranking the pupils validly and reliably were thus judged by comparing the correlations they both obtained against the Criterion Test.

(*b*) *Judgment of O-grade standards*

The criterion of standards, or of the appropriate levels in the rank order above which to award bands A, B, C, D, and E, had to be the judgment of the O-grade examiners in assigning ranges and bands to pupils in the examination: the researchers could not guarantee to set standards for the Criterion Test equivalent to those of the O-grade. (In the event, the Criterion Test as set and marked was probably a little more severe than the 1977 O-grade examination.) Accordingly performance in the O-grade was the basis of comparison when standards were considered, and the procedure adopted was to compare Folio and O-grade pass rates at each band before and after the scaling of Folio marks against those scored by the same pupils in the external examination.

---

[1] No reliability coefficient was calculated for the Criterion Test because it was thought that, since each paper sought to test at least some different skills, a split-half method of calculating one, or any variation of it, was inappropriate.

III. Note on Terminology: Marks, Standardising, Scaling, Ranges and Bands

Marks may be 'raw', 'standardised' or 'scaled'.

Raw marks are those actually awarded by the markers to scripts. 'Standardisation' usually means re-expressing raw marks in some standard way: e.g., all O-grade raw marks in all subjects are standardised to a pass mark of 50% and a standard deviation of 20. The term 'scaling' normally refers to the process by which a set of raw marks is adjusted so that its mean and standard deviation are the same as that of another set of marks for the same pupils – e.g., raw marks awarded in the school Folio assessment and in the Criterion Test could be scaled against the pupils' marks in the O-grade.

Some confusion is possible, because SCEEB officials and examiners commonly use the word 'scaling' in reference to *both* the processes described in the previous paragraph, while they use 'standardisation' to refer to a different process, the one by which each marker's consistency and accuracy is checked by the Principal Examiner and his team.

In this report, the following definitions obtain, unless the text specifically indicates otherwise.

Raw marks: marks actually awarded to scripts by markers.
Standardisation: the re-expression of raw marks so that the pass mark decided on by the Principal Examiner becomes 50% and the standard deviation of the marks becomes 20. (This is often called 'scaling' by SCEEB.)
Scaling: the adjustment of a set of marks so that its mean and standard deviation are the same as those of another set of marks for the same pupils.
Marker-standardisation: the process by which the Principal Examiner and his colleagues check the consistency and accuracy of individual markers' work and make necessary adjustments.
Ranges and bands: standardised categorisations of marks employed by SCEEB in reporting examination results. The relationship between SCEEB standardised marks and ranges and bands is shown in Table 3.1.

TABLE 3.1: MARKS, RANGES AND BANDS

Standardised mark	*Range*	*Band*
90-100	1	A
85-89	2	
80-84	3	
75-79	4	
70-74	5	
65-69	6	B
60-64	7	
55-59	8	C
50-54	9	
45-49	10	D
40-44	11	
35-39	12	E
30-34	13	
0-29	14	No award 'F'

Note: In the SCRE project, for convenience, 'No award' has been called band 'F'.

CHAPTER IV

FOLIO AND O-GRADE: RANKING AND PASS RATES

I. Ranking - Comparison of both Folio and O-grade with Criterion Test

(*a*) *Overall Impression*
Pearson Product-moment correlation coefficients were calculated for Folio/Criterion Test, O-grade/Criterion Test, and Folio/O-grade.

The three coefficients could be obtained for 11 schools; of the remaining three schools, two provided no folio assessments, so that only the O-grade/Criterion Test coefficient was found for them; and one did not set the Criterion Test and had to be left out of the calculations. Some figures are available, therefore, for 13 schools, all figures for 11.

In order to give a general impression of the validity of each of the three assessment methods by comparison with the other two, Table 4.1 shows the coefficients for all the pupils in the 13 schools for whom data had been obtained. It should be noted (1) that the three coefficients are not strictly comparable because they were calculated for three slightly different populations, (2) that the quality of assessments in individual schools and classes cannot be judged from these correlations. *In very general terms however, Folio Assessment, O-grade and the Criterion Test can be said to correlate with one another satisfactorily.*

TABLE 4.1: PEARSON CORRELATION COEFFICIENTS

r Folio/Criterion Test	*O-grade/Criterion Test*	*Folio/O-grade*
·77 (1567 pupils)	·82 (1834 pupils)	·71 (1529 pupils)

(*b*) *Folio Assessment on a Whole School Basis*
Judgments about the relative merits of Folio Assessment and O-grade could be made only after consideration of the correlations by school and by class which are given in Tables 4.2 and 4.3.

Interpretation of correlation coefficients is not straightforward: the value of the coefficient depends, not only on the quality of the assessments, but also on the size of the sample and the distribution of marks within it. Thus a school presenting for O-grade and Folio Assessment only those pupils likely to obtain band C or better is likely to show a lower correlation coefficient than one without restrictions on the type of pupils sitting the examination, the quality of assessment being equal in both. One cannot, therefore, read *down* Table 4.2 reaching conclusions about the relative quality of schools' Folio Assessments simply by comparing the coefficients obtained by different schools. Caution is needed. The Notes columns of the Tables provide information of which account should be taken in evaluating some of the correlations. It is however, legitimate to make comparisons *across* each row in the Tables, since the pupils involved are the same (or virtually so). The size of the Folio/Criterion Test coefficient relative to that for O-grade/Criterion Test indicates the degree by which Folio Assessment was better or worse than the O-grade at ranking the pupils on those skills measured by the Criterion Test.

TABLE 4.2: PEARSON CORRELATION COEFFICIENTS: SCHOOLS ASSESSING ON WHOLE-SCHOOL BASIS

School	*r Folio/ Criterion Test*	*r O-grade/ Criterion Test*	*r Folio/ O-grade*	*Notes*
11	·74 (144 pupils)	·69 (143 pupils)	·68 (143 pupils)	(1) Folio assessment made in ranges 1-14, not in % marks – so ranking was less refined than elsewhere. (2) Truncated: poorer pupils not entered.
12	no folio	·73 (99)	no folio	Part of Criterion Test taken in class: some pupils therefore failed to complete it.
13	·68 (163)	·85 (165)	·61 (163) ·71 (262)	Part of Criterion Test taken in class: 165 pupils only completed the full test, though 262 sat O-grade and received Folio assessment.
14	·83 (227)	·81 (218)	·74 (196)	Truncated at top: best pupils by-pass O-grade.
15	·84 (155)	·84 (154)	·81 (154)	
21	·86 (94)	·82 (93)	·83 (93)	Truncated at bottom: poorer O-grade pupils excluded from project by school.
22	·84 (204)	·85 (209)	·79 (204)	
23	·84 (81)	·87 (89)	·79 (81)	
24	·89 (72)	·88 (70)	·84 (70)	

Note: School 12 dropped out of the Folio scheme early in the session.

The picture presented by seven of the remaining eight schools carrying out whole-school assessment is a satisfactory one. Six of the Folio/Criterion Test coefficients are better than ·80. The lower value for school 11 (·74) may be partly explained by two factors: (a) the Folio assessment was reported to SCRE in discrete ranges 1-14 and not in continuous marks 1-100; (b) the distribution of ability in the sample was restricted by a policy of not entering poorer pupils for the examination and by the relative lack of exceptionally high achievers. (Of 38 pupils in school 11 obtaining O-grade band A, 11 had ranges 1, 2 or 3, as against 14 out of 30 in school 12 and 16 out of 31 in school 15, two nearby schools.)

The 't' test for significance in the difference between two correlations which are also correlated with each other was applied.

In seven of the eight cases of whole-school assessment there is no significant difference between the Folio/Criterion Test and the O-grade/Criterion Test cor-

relations. These schools can be said to have ranked their pupils satisfactorily by comparison with an independent criterion and to have done so as well as the O-grade examination did by comparison with the same criterion.

In the case of school 13, the superiority of the O-grade/Criterion Test correlation over the Folio/Criterion Test correlation is highly significant. Possible reasons for the distinctly poorer Folio/Criterion Test correlation in this school are suggested below (see page 35).

(*c*) *Folio Assessment on a Class Basis*
Tables 4.3 and 4.4 show the correlations for the Strathclyde schools, which had presented Folio Assessments carried out by teachers for their own classes.

TABLE 4.3: PEARSON CORRELATION COEFFICIENTS: SCHOOLS ASSESSING BY CLASS – SCHOOL CORRELATIONS

School	*r Folio/ Criterion Test*	*r O-grade/ Criterion Test*	*r Folio/ O-grade*	*Notes*
32	·83 (268 pupils)	·81 (270 pupils)	·71 (268 pupils)	
33	·55 (45)	·84 (126) ·71 (45)	·39 (45)	Folio assessments for two classes only.
34	no folio	·86 (85)	no folio	Many pupils did only part of the Criterion Test: 85=only about two thirds of the presentation group.
35	·63 (114)	·84 (113)	·59 (112)	Some classes not included in project, because of teachers' unwillingness, but remainder cover a wide range of ability.

School 34 was unable to provide folio marks, and school 33 could provide only partial data. (The difference between ·55 for Folio/Criterion Test and ·71 for O-grade/Criterion Test for 45 pupils in this school is not significant.) Of the two Strathclyde schools providing full data, school 32 has no significant difference between the correlations and school 35 has a Folio/Criterion Test correlation significantly lower than that for O-grade/Criterion Test.

TABLE 4.4: PEARSON CORRELATION COEFFICIENTS: SCHOOLS ASSESSING BY CLASS – CLASS CORRELATIONS

School	*Class*	*r Folio/ Criterion Test*	*r O-grade Criterion Test*	*r Folio/ O-grade*	*Notes*
32 First half of alphabet, roughly set by previous assessments (in English)	a1	·75 (25)	·59 (26)	·63 (25)	Ranges obtained in O-grade =1-9.
	a2	·81 (24)	·66 (24)	·58 (24)	Ranges obtained in O-grade = 1-13. – Folio assessed in *ranges* not marks.
	a3	·81 (27)	·48 (27)	·32 (27)	Ranges obtained in O-grade =4-13.
	a4	·35 (24)	·72 (24)	·27 (24)	Ranges obtained in O-grade = 6-14. – Folio assessed in ranges.
	a5	·48 (24)	·63 (25)	·30 (24)	Ranges obtained in O-grade =8-14.
	a6	·78 (13)	·75 (13)	·78 (13)	Ranges obtained in O-grade =10-14.
Second half of alphabet roughly set by previous assessments (in English)	b1	·64 (26)	·60 (26)	·47 (25)	Ranges obtained in O-grade =1-10.
	b2	·57 (26)	·39 (26)	·41 (24)	Ranges obtained in O-grade =3-12.
	b3	·75 (26)	·52 (26)	·63 (27)	Ranges obtained in O-grade =3-13.
	b4	·59 (25)	·46 (25)	·39 (24)	Ranges obtained in O-grade =5-14.
	b5	·77 (20)	·72 (20)	·74 (24)	Ranges obtained in O-grade =6-14.
	b6	·92 (8)	·81 (8)	·78 (13)	Ranges obtained in O-grade =11-14. – Pupils widely scattered within marks range 0-44.
33 Classes set by previous English assessments	1	no folio	·65 (27)	no folio	Ranges obtained in O-grade =1-9.
	2	·39 (5)	·63 (25)	·81 (5)	Ranges obtained in O-grade =1-10.
	3	·55 (21)	·50 (21)	·03 (21)	Ranges obtained in O-grade =2-13.
	4	·45 (19)	·61 (19)	·33 (19)	Ranges obtained in O-grade =6-14. – Folio assessed in ranges.
	5	no folio	·73 (18)	no folio	Ranges obtained in O-grade =6-14. – Folio assessed in ranges.
	6	no folio	·49 (16)	no folio	Ranges obtained in O-grade =7-14.
35 Classes set by previous English assessments	1	·49 (27)	·50 (27)	·36 (27)	Ranges obtained in O-grade =1-5. – (+1 range 7)
	4	·75 (24)	·47 (22)	·56 (22)	Ranges obtained in O-grade =1-14.
	5	·61 (20)	·77 (20)	·49 (20)	Ranges obtained in O-grade =3-14.
	7	·71 (21)	·69 (21)	·65 (21)	Ranges obtained in O-grade =3-11.
	8	·46 (12)	·76 (12)	·64 (12)	Ranges obtained in O-grade =5-13.
	9	·82 (10)	·83 (11)	·64 (10)	Ranges obtained in O-grade =5-14.

Since three of the five schools committed to assessment by the class teacher did not provide adequate data, judgments about the effectiveness of this method are not very firmly based. There are, however, some indications that class teacher assessments are not adequate as a basis for public certification. Difficulties in moderating class assessments, either statistically or by marking sample scripts, are obvious. Some indications of 'class effects' even in the 'whole-school' assessment group also lead to the conclusion that team assessment is desirable to counteract teachers' idiosyncrasies in setting tasks or marking. The common practice of setting S4 English classes by 'ability groups' can cause class assessments to be more inaccurate by comparison with a criterion than they might have been: teachers may rank pupils very well within their own class but be influenced in making awards by the supposed 'quality' of the set, so that, for instance they award only band A to the top class and nothing above band E or D to the bottom one. This may result in a very unsatisfactory rank order for the whole school, since the assessments on which setting was based in the first place were probably not accurate enough to assign pupils with certainty to different ability groups. Though there may be faults in O-grade ranking too, the bands achieved in the examination by pupils in a particular class are a fair indication of the breadth of achievement in it. It can be seen from the 'Notes' column of Table 4.4 that 'ability' sets in English may very easily contain pupils at almost all levels of O-grade achievement.

The effects of previous setting can be seen in two of the 'by class' schools. Table 4.4 gives the correlation coefficients by class for this group of schools, and shows that in schools 32 and 35, especially the former, class teachers did a creditable job in ranking their own classes. Indeed, when the class correlation coefficients for school 32 were combined and the appropriate statistical test was carried out, it was found that the Folio/Criterion Test correlations were significantly better than the O-grade/Criterion Test ones. The teachers here had, then, ranked their separate classes better than the O-grade did, if Criterion Test results are taken as the criterion. This does not mean, however, that they produced a *school* rank order which was superior: in fact, Table 4.3 shows that there was no significant difference between *school* Folio and O-grade correlations with the Criterion Test, and the school rank order for school 35, where the teachers also fairly successfully ranked their own classes, was actually significantly less satisfactory than the O-grade one. The reason for this apparent contradiction is that the means by which classes had previously been set had not in fact separated pupils into homogeneous ability groups. In school 32 the top and bottom sections had been fairly well identified, but in the middle groups there were pupils obtaining all O-grade bands, A to 'F'; in school 35 this happened in all the classes for which we have data, except the top section. (See Table 4.4, 'Notes' column.) Since classes were, in effect, of mixed ability, it was easier for teachers to rank within them than if they had been genuine ability groupings. For the same reason, there was a large overlap at both ends between a class rank order and that of the classes above and below it; and this resulted in a less satisfactory *school* correlation with the Criterion than might have been expected from apparently quite good within-class ranking.

The principal argument in favour of class teacher assessment as opposed to team assessment is that it is less time-consuming. The researchers believe, however, that the extra time required for departmental standardising is very well spent and that the advantages accruing from it are significant. The very good class teacher assessment in school 32 does not belie this belief, because the teachers in that school did in fact approach the whole project very much as a team and there was much intra-departmental consultation and discussion, even though each carried out his or her own class assessments.

It is possible that class assessments made an important contribution to the relatively poor internal assessments in school 13 (see Table 4.2). There was some evidence from the Trial Markings and the moderation procedure in this school that the department had been unable to keep to the instruction to work as a team, each marking the work of a cross-section of the year group.

It is the conclusion of the researchers that class teacher assessments should be rejected as a means of public certification because of the difficulties of standardising them, difficulties which, though present, are more easily overcome for whole-school assessments.

(*d*) *Some Possible Causes of the Lower Correlations*

Before leaving the consideration of ranking it may be helpful to call to mind factors affecting correlations and, in particular, to try to identify possible reasons for the less satisfactory performance of schools 13 and 35.

The correlations between the three sets of assessment results could have been affected by some or all of six factors.

(1) Variability in the performance of the pupils on different occasions.
(2) The character of the group of pupils concerned.
(3) Lack of effective discrimination in Folio and the two examinations prior to scaling.
(4) Marker inconsistency.
(5) The influence of the task set and of choice of tasks, especially in assessing interpretation.
(6) Organisational difficulties.

The effects of some of these influences on all the schools is discussed in detail in Chapter VII.

There are no reasons for thinking that schools 13 and 35 are different from the others with regard to numbers 1-3, except that they did happen to be the schools obtaining, respectively, the lowest and highest mean marks in the O-grade examination.

All the schools probably suffered from a degree of inconsistency in the marking (Chapter VII shows how much is likely); there may have been more inconsistency in schools 13 and 35 than elsewhere, but, if so, the concern would be to try to find causes for it. The deleterious influence of class assessments in school 35, and possibly also in school 13, has already been mentioned. It is closely related to the problems caused by variations in the level of difficulty in the tasks set for pupils – in particular, different class teachers probably set more and less difficult interpretation tests. If this drawback was combined with a failure to involve all the staff in the Trial Marking exercises, a school's marks would be more vulnerable to inconsistency among the markers than others. Teachers in school 35 had, of course, been asked to work independently, though they did all take part in the Trial Marking exercises. There were indications that it was more difficult to engage everybody fully in the project in school 13, and, although it is not possible to account for a low correlation solely in terms of organisational and administrative difficulties, they probably did play a part in lowering the Folio/Criterion Test correlation there. Only a proportion of the pupils completed the Criterion Test there because parts of it were set in class and not in the same standard examination conditions as in the other schools. A practical difficulty, therefore, considerably reduced the value of the reference test. There were also indications in the data that Folio Assessment there may have been less co-ordinated than in more successful schools. A number of characteristics suggesting 'by-class' assessment were noted, though a 'whole-school' approach was required. Communication within the department may have been difficult: in responding to the questionnaire at the end of the project, three teachers reported that they had no knowledge of a moderator's visit to the school to discuss the appropriateness for O-grade of work in progress, though one did take place.

Comments on organisational problems, which should be linked with the report on practical problems in Chapter IX, are only impressionistic, since it was no part of the remit of the project to judge the quality of teaching or management in schools. It did seem, however, that a number of the difficulties which were met in the course of the experiment in all schools were essentially management problems. How to allocate very limited marking time? How to find time for departmental discussion of the scheme? How to obtain co-operation from reluctant members of staff? How to maintain folders in good order? And so on.

For successful Folio Assessment it is probably necessary to have a member of the department who is committed to the method and has the time, energy, knowledge and authority to encourage others, arrange meetings, give advice on standards within the school, impose deadlines and insist on adherence to instructions and the full involvement of all in Trial Marking and Moderation procedures. It is, perhaps, a role for a very capable Assistant Principal Teacher.

One remaining factor characterised school 35: several teachers there entered 'estimate' marks for pupils who had failed to complete the minimum requirements of the project or whose performance in the Folio work did not match the teachers' view of their 'ability'. There may have been some confusion in the minds of these teachers as to what exactly they were assessing.

It would be necessary in a real Folio Assessment scheme to make absolutely clear the instruction to assess only the work actually produced, since estimates of potentiality are even more likely to be unreliable than judgments about actual performance in English.

II. Pass Rates - Comparison of Folio Standards with O-grade Standards

Schools' accuracy in assigning O-grade bands may be judged from a comparison of pass rates at each band between Folio Assessment and O-grade. A similar comparison between Folio and Criterion Test would have been possible, but would have been superfluous as far as consideration of *standards* is concerned: the only valid basis for *standards* available was the awards made by the O-grade examiners; so bands A, B, C, D and E on the Criterion Test could have been awarded only after scaling its results against the O-grade results.

The O-grade and Folio pass rates, before scaling the latter, are given in Tables 4.5 and 4.6 for 1791 pupils, all those in the 11 schools for whom both Folio and O-grade marks were obtained.

From the 'Folio (raw)' column can easily be seen the typical tendency of internal assessment to bunch the marks near, and especially just above, the C pass mark, along with the ensuing failure to award enough As, Es and 'F's.

Raw Folio marks clearly do not produce pass rates at each band comparable to those of the O-grade examination.

The simple comparison of pass rates set out in Table 4.6 is not, however, fair, and Folio assessment cannot be condemned on the basis of it. It is necessary to compare Folio with O-grade after the former has been subjected to moderation to make it conform to O-grade standards. The effects of moderation are described in the following chapter.

TABLE 4.5: OVERALL PASS RATES AT EACH BAND: 1791 PUPILS IN 11 SCHOOLS

Band	*O-grade*	*Folio (raw)*
A	22·3%	12·2%
B	17·4%	20·4%
C	21·4%	37·4%
D	16·4%	20·3%
E	11·6%	7·6%
'F'	10·9%	2·1%

TABLE 4.6: PASS RATES AT EACH BAND, BY SCHOOL

School	*Band*	*O-grade*	*Folio (raw)*
11 (149 pupils)	A	26%	14%
	B	20%	20%
	C	27%	44%
	D	17%	18%
	E	8%	4%
	'F'	3%	0%
13 (263 pupils)	A	17%	8%
	B	12%	21%
	C	19%	37%
	D	14%	19%
	E	13%	14%
	'F'	25%	2%
14 (217 pupils)	A	24%	5%
	B	18%	24%
	C	18%	45%
	D	15%	20%
	E	13%	5%
	'F'	12%	1%
15 (163 pupils)	A	20%	20%
	B	17%	10%
	C	19%	31%
	D	19%	30%
	E	15%	9%
	'F'	10%	0%
21 (110 pupils)	A	27%	25%
	B	15%	25%
	C	26%	40%
	D	22%	8%
	E	7%	1%
	'F'	3%	1%
22 (236 pupils)	A	27%	11%
	B	24%	16%
	C	17%	35%
	D	17%	25%
	E	8%	8%
	'F'	7%	4%
23 (109 pupils)	A	19%	10%
	B	14%	19%
	C	24%	29%
	D	15%	22%
	E	14%	14%
	'F'	14%	6%

TABLE 4.6 (contd)

School	*Band*	*O-grade*	*Folio (raw)*
24 (73 pupils)	A	22%	7%
	B	29%	23%
	C	15%	30%
	D	15%	15%
	E	4%	19%
	'F'	15%	6%
32 (281 pupils)	A	17%	14%
	B	18%	24%
	C	22%	37%
	D	17%	16%
	E	15%	6%
	'F'	11%	3%
33 (59 pupils)	A	12%	3%
	B	9%	19%
	C	44%	53%
	D	13%	23%
	E	17%	2%
	'F'	5%	0%
35 (131 pupils)	A	34%	18%
	B	15%	24%
	C	24%	32%
	D	15%	24%
	E	8%	2%
	'F'	4%	0%

CHAPTER V

MODERATION

Three types of moderation device were employed:

I 'Trial Marking' by all the schools of sample scripts provided to try to establish common standards.

II Marking of a representative sample of each school's folios by external moderators.

III Scaling of each school's folio assessments against its performance in the O-grade examinations.

The function of moderation in the SCRE scheme was to ensure that internal pass rates were as near as possible to those the schools would have obtained in the external examination and to ensure as much fairness to every individual candidate as possible.

Of the three approaches to moderation attempted, the effects of the visiting moderator scheme and of scaling against O-grade performance could be gauged from the statistical evidence, but that of the Trial Marking exercises could not; teachers' response to these exercises, however, and the evidence from other exam boards of their value as in-service training[1] indicate that they should be included in any internal assessment scheme.

Appendices 5 and 6 contain detailed descriptions of both Trial Marking arrangements and the Visiting Moderator scheme. The concern here is only with the *effectiveness* of the Visiting Moderator and of Scaling for ensuring comparability between Folio and O-grade.

I. Visiting Moderator

Apart from school 33, where judgments about the sample folios for moderation were unreliable because several of the folders supplied were incomplete, all the schools passed the three moderation tests (which were those developed for CSE, described in the Schools Council Examination Bulletin No. 5 and used by TWYLREB (The West Yorkshire and Lindsey Regional Examination Board). The teachers and the moderators were in reasonable agreement as to ranking, standards and the discriminations made among the pupils.

All the school awards were therefore vindicated by the visiting moderators, and the discrepancy in pass rates (see Tables 4.5 and 4.6) was not corrected by the moderators.

The discrepancy in pass rates was caused by three factors which the three moderation tests aim to check:

1. disagreement on the size of the distribution of marks (discrimination);
2. disagreement on standard, or excessive severity or leniency; the Schools Council test accepts as satisfactory discrepancies between moderator and school of up to half a grade, or, in Scottish terms, one range;
3. disagreement on ranking, which, if it is significant and consistent, probably means that moderator and teachers are valuing different things in the scripts.

In which of the three areas did the moderators' failure occur?

---

[1] See, e.g. Cohen and Deale, (Schools Council, 1977), and Chapter X of this report.

1. *Discrimination: failure to overcome 'bunching'*

The failure of the moderators to detect 'bunching' is not surprising. Firstly, the size of a sample which can be re-marked in a realistic time makes bunching harder for the school to illustrate. The schools were asked to choose the sample so that the number of moderated pupils in each band was proportionate to the whole school numbers in each band, but this may in fact have been too time-consuming an inconvenience to be kept to. Secondly, and probably more significantly, both teachers and moderators were marking with the O-grade standardised marks pattern in mind; i.e. 70 plus=A, 60 plus=B, and so on. Markers of the O-grade papers also mark with this categorisation in view, and they do in fact produce a similarly bunched distribution to that obtained in the Folio Assessment (as can be seen from the superimposition of the histogram of the Folio distribution of raw marks on that of the O-grade distribution of raw marks, which is contained in Appendix 10). A standard deviation of about 12·5 is normal for O-grade English raw marks, and that for the pupils involved in the project was actually 11·78. A distribution of this sort means that 3 or 3·5 raw marks are equivalent to one O-grade range, and 6 or 7 raw marks to one band.

It is not the O-grade English markers and examiners, nor, therefore, the moderators of an internal assessment scheme who effect the apparently clear discrimination of pupils into bands A to 'F', but the statistical standardising procedure, which stretches out raw mark ranges of 3 or 3·5 marks and bands of 6 or 7 to standardised ranges of 5 marks and bands of 10. Most of the discrepancy in pass rates between Folio and O-grade in Table 4.5 is due to the fact that raw Folio marks have been compared with standardised O-grade results, and the differences are principally in the distribution of marks rather than in the means.

It might be thought that the standardising procedure gives O-grade English results a spurious appearance of fine discrimination; and, when inter-marker inconsistency is considered (see Chapter VII) the reliability of the discriminations made by the markers is in rather more doubt than is normally the case in the mind of the public, including teachers and markers themselves. Since small errors in assigning pupils to their proper places in the rank order may be exaggerated by the standardising procedure, some individual pupils may obtain awards quite different from those intended by the marker. (The effects on individual pupils of scaling Folio marks are discussed below, pages 44-49.)

It is clear that, if the aim is to obtain Folio awards distributed in the same way as O-grade awards, some form of the SCEEB statistical standardising procedure must be applied to the Folio marks: the Visiting Moderator will not be effective for this purpose. The standardising procedures, however, impose an appearance of finer discrimination among candidates than is actually achieved by the markers.

2. *Severity/leniency*

It would be possible for a Visiting Moderator to check the severity or leniency of schools' raw marks, leaving the standardisation of spread of marks to be effected by the same procedure which at present applies to the O-grade marks, viz., the distribution of awards for the whole examination is adjusted so that a given standard deviation is obtained, and the standard deviation for each school is then calculated.

Did the Visiting Moderator scheme in the SCRE project in fact show that it could satisfactorily correct severity and leniency?

Since the Moderator was representing the O-grade examination, the difference between the Folio raw marks mean and the Folio mean after scaling against the

O-grade marks can be taken as a measure of each school's severity or leniency relative to the O-grade. These differences were:

School 11	+0·95	Raw Folio severe
School 13	−4·57	Raw Folio lenient
School 14	−0·68	Raw Folio lenient
School 15	−2·65	Raw Folio lenient
School 21	−1·82	Raw Folio lenient
School 22	+6·50	Raw Folio severe (beyond normally acceptable limits)
School 23	−0·40	Raw Folio lenient
School 24	+4·05	Raw Folio severe
School 32	−3·51	Raw Folio lenient
School 33	−2·34	Raw Folio lenient
School 35	+3·90	Raw Folio severe

While there may be statistically significant differences between means at the 95% confidence limits, it is normally taken that differences smaller than half a grade in CSE terms, or one range in SCEEB terms, are not educationally meaningful[1]. A difference of 5 percentage marks, which is significant at the 99% confidence limits, would mean that a school had marked severely or leniently to the extent that a noticeable proportion of its candidates had received awards the same as those which the O-grade was making to other pupils whose work was of a standard which examiners could recognise as being clearly different. Accordingly, the moderation test employed accepted as satisfactory differences between the moderator's mean and that of the school of up to one range, or 5 percentage marks.

In fact the test for severity/leniency was passed by all the schools, except school 33. When the schools' raw and scaled means are compared, however, it is found that school 33, though lenient, is well within the acceptable limits, while school 22, where the school assessments had been vindicated by the moderator, was, in fact, unacceptably severe.

In addition to these two failures, the SCRE moderation suffered from the confusion about the real size of 'one range', which was referred to above in the discussion of distribution of marks. While teachers, O-grade markers, and moderators made their assessments with the idea in mind that one range equals five marks, they actually bunched the distribution so much that, in reality, the size of a range was about three or three-and-a-half marks. By this criterion schools 13, 24, 35 and (just) 32 were also severe or lenient by more than one range, though the moderation tests were not designed to function with as fine a degree of accuracy as would be necessary to check discrepancies of only three marks – indeed, it is very doubtful if discrepancies of that order are significant educationally (as opposed to statistically).

It is clear that the actual narrow distribution of raw percentage marks, combined with the belief on the part of teachers, markers, and moderators, that ranges are wider than they really are, causes serious difficulties for effective moderation of standard as well as of spread of marks.

The SCRE moderators, despite their experience as O-grade markers and their having satisfactorily passed the relevant tests at their own Trial Marking, were not able to correct severity or leniency adequately.

[1] For confirmation of these assertions, see Schools Council Examination Bulletin 36, 'Mode Comparability in the CSE', by Bloomfield, Dobby and Duckworth. (1977)

This failure may have been due to their inevitable lack of experience as moderators, to the pressure to complete unfamiliar work in a short time, or to the unrepresentativeness of the schools' sample scripts. The task is made harder because the bunching of marks and subsequent standardising cause the published O-grade results to make discriminations among pupils obtaining contiguous ranges which are too fine to expect moderators to make them reliably.

The conclusion of the researchers is that the SCRE moderation was not successful enough to ensure comparability of O-grade English standards across schools, but with the significant note that reliable classification into ranges of 3 or 3·5 raw marks width is almost impossible in English.

3. *Disagreement on ranking*

The TWYLREB test for 'conformity', or, in effect, agreement on ranking, is designed in such a way that, if it is passed, the correlation between the school marks and the moderators' for 20 sample scripts is not less than about ·70. (See Schools Council, 1965, page 25.) This does not mean, of course, that the correlation for the moderators' and the teachers' marks for the whole school would be the same, but it should mean that schools passing the test are in reasonable agreement on ranking with the external examination, for which, in the case of the SCRE project, the moderator was a possible substitute. In fact, although all schools passed this test, it has already been shown in Tables 4.2 and 4.3, that schools 13 and 35 obtained Folio/O-grade correlation coefficients of ·61 and ·59 respectively, figures which would not normally be regarded as indicating acceptable comparability of ranking between two assessments.

The Visiting Moderator procedure therefore failed also to identify the two internal assessments which were relatively unsatisfactory with regard to ranking.

II. Scaling

(*a*) *Effect on Pass Rates*

Folio marks might have been scaled against raw O-grade marks so that each school's mean and distribution of marks was the same as those it obtained in the O-grade before standardisation of the latter. It would then have been necessary to apply the SCEEB standardising procedure to the whole group of Folio pupils, in the same way as it is applied to the whole population of O-grade candidates. The method employed by the researchers was, however, to apply scaling and a form of standardising to the Folio marks simultaneously by scaling them against the O-grade standardised marks.

Leaving aside the possible disadvantages for the accuracy of individual pupils' awards, it is clear from Tables 5.1 and 5.2 that pass rates at each band were made much more comparable by scaling Folio marks against O-grade standardised marks (though a slight bulge still remained in Folio bands C and D).

TABLE 5.1: OVERALL PASS RATES AT EACH BAND: 1791 PUPILS IN 11 SCHOOLS

Band	*O-grade*	*Folio (raw)*	*Folio (scaled against O-grade)*
A	22·3%	12·2%	21·7%
B	17·4%	20·4%	15·4%
C	21·4%	37·4%	22·8%
D	16·4%	20·3%	20·4%
E	11·6%	7·6%	10·0%
'F'	10·9%	2·1%	9·7%

TABLE 5.2: PASS RATES AT EACH BAND, BY SCHOOL

School	*Band*	*O-grade*	*Folio (raw)*	*Folio (scaled)*
11 (149 pupils)	A	26%	14%	22%
	B	20%	20%	19%
	C	27%	44%	44%*
	D	17%	18%	12%
	E	8%	4%	6%
	'F'	3%	0%	3%
				*Note: Scaling had no effect on the number of C awards because Folio Assessment was made in ranges, not marks.
13 (263 pupils)	A	17%	8%	17%
	B	12%	21%	11%
	C	19%	37%	17%
	D	14%	19%	22%
	E	13%	14%	12%
	'F'	25%	2%	21%
14 (217 pupils)	A	24%	5%	20%
	B	18%	24%	15%
	C	18%	45%	22%
	D	15%	20%	22%
	E	13%	5%	10%
	'F'	12%	1%	11%
15 (163 pupils)	A	20%	20%	21%
	B	17%	10%	9%
	C	19%	31%	21%
	D	19%	30%	22%
	E	15%	9%	21%
	'F'	10%	0%	6%
21 (110 pupils)	A	27%	25%	26%
	B	15%	25%	14%
	C	26%	40%	30%
	D	22%	8%	23%
	E	7%	1%	5%
	'F'	3%	1%	2%
22 (236 pupils)	A	27%	11%	27%
	B	24%	16%	24%
	C	17%	35%	22%
	D	17%	25%	16%
	E	8%	8%	6%
	'F'	7%	4%	5%
23 (109 pupils)	A	19%	10%	17%
	B	14%	19%	17%
	C	24%	29%	19%
	D	15%	22%	24%
	E	14%	14%	6%
	'F'	14%	6%	16%
24 (73 pupils)	A	22%	7%	26%
	B	29%	23%	26%
	C	15%	30%	10%
	D	15%	15%	12%
	E	4%	19%	14%
	'F'	15%	6%	12%

TABLE 5.2 (contd)

School	*Band*	*O-grade*	*Folio (raw)*	*Folio (scaled)*
32 (281 pupils)	A	17%	14%	19%
	B	18%	24%	15%
	C	22%	37%	21%
	D	17%	16%	24%
	E	15%	6%	10%
	'F'	11%	3%	11%
33 (59 pupils)	A	12%	3%	12%
	B	9%	19%	17%
	C	44%	53%	29%
	D	13%	23%	25%
	E	17%	2%	14%
	'F'	5%	0%	3%
35 (131 pupils)	A	34%	18%	34%
	B	15%	24%	34%
	C	24%	32%	22%
	D	15%	24%	20%
	E	8%	2%	8%
	'F'	4%	0%	2%

(*b*) *Effect on individual awards*

Broadly speaking, therefore, scaling corrected the mean and distribution of internal assessment marks to match the O-grade ones. What were its effects on individual school pupils?

It is helpful to consider the effect of the discrepancy in pass rates at each band before and after scaling on the 'match' between Folio and O-grade awards, i.e., on the numbers of pupils receiving the same range, plus or minus one range, etc., in the two assessments.

Tables 5.3 and 5.4 give the overall figures before and after scaling.[1]

TABLE 5.3: 'MATCH' OF AWARDS: FOLIO (RAW) – O-GRADE
(1791 PUPILS IN 11 SCHOOLS)

No. of Ranges misplaced	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9
Frequency	1	3	16	58	68	132	221	264	289	240	211	149	84	34	14	4	2	1
%			1%	3%	4%	7%	12%	15%	16%	13%	12%	8%	5%	2%	1%			

Folio more lenient ← || → Folio more severe

Same range = 16%
Within ±1 range = 44%
±2 ranges = 68%
±3 ranges = 83%

(percentages are approximate)

Assessed more severely by Folio = 41%

Assessed more leniently by Folio = 42%

---

[1] Appendix 12 contains figures for each school separately.

TABLE 5.4: 'MATCH' OF AWARDS: FOLIO (SCALED AGAINST O-GRADE) – O-GRADE (1791 PUPILS IN 11 SCHOOLS)

No. of ranges mis-placed	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9	10
Fre-quency	1	1	4	9	17	44	75	128	175	251	377	266	176	123	69	46	20	4	3	0	2
%				1%	1%	2%	4%	7%	10%	14%	21%	15%	10%	7%	4%	3%	1%				

Folio more lenient ← → Folio more severe

Same range = 21%
Within ±1 range = 50%
Within ±2 ranges = 70%

Assessed more severely by Folio = 30%

Assessed more leniently by Folio = 39%

(percentages are approximate)

Scaling against O-grade standardised marks improved the percentage of pupils within ±1 range by 6% (to 50%) and that of pupils within ±2 ranges by 2% (to 70%).[1]

Some extreme discrepancies between Folio and O-grade ranges were, however, exaggerated by the scaling process, so that the greatest differences between Folio and O-grade ranges after scaling became +10 (2 cases) and −10 (1 case), whereas they had previously been +9 and −8 (1 case each).

When each school was considered individually, it was discovered that the match of awards of ranges between Folio and O-grade was not always improved by the scaling, and in some cases was less good after it. There were also several cases among the schools where the greatest discrepancy was larger after scaling than before.

---

[1] **A Note on School Examinations**
An incidental but notable point is that the percentages of pupils obtaining the same award within ± 1 and 2 ranges on Folio and O-grade reported in Table 5.4 may be compared with those reported by SCEEB in *An Investigation into the Comparability of School Estimates and Examination Performance* (SCEEB, 1974) when school examination marks were matched against O-grade marks. The percentage of pupils assigned the same range then (before scaling) was 19%; 49% were within ± 1 range, and 73% within ± 2 ranges, as compared with 16%, 44%, and 68% for folio assessment. It is not possible to conclude from these figures that school examinations are better than Folio Assessment by comparison with the O-grade examination, since there may be significant differences between the character of the two study populations, the SCEEB investigation having been carried out before the raising of the school leaving age had encouraged many less able pupils to take the examination. School examinations are, however, more like the O-grade than Folio assessment, and this is true in particular of the testing of interpretation, which the SCRE project found to be unsatisfactory in Folio Assessment. If internal assessment were intended only to duplicate the function of external examinations, a school examination would certainly be as satisfactory a means of fulfilling the requirement as Folio Assessment, given adequate 'Trial Marking' experience for the teachers and some moderation of the level of difficulty of the examination set.

The following table shows the percentage of pupils in each school within ±1 range on Folio and O-grade before and after scaling of the former. It also shows the effect of scaling on the mean and standard deviation of each school's Folio marks. The amount of discrepancy remaining after scaling is that due to the influence of the various factors affecting ranking such as variable performance by the pupils and unreliability in both Folio and O-grade: the correlation coefficients given in the table for each school, though calculated for very slightly different numbers, are indications of the amount of agreement on ranking between Folio Assessment and O-grade, so that it is possible to see approximately how many pupils in each school received different awards (after scaling) in Folio and O-grade, when the correlations were of given sizes.

TABLE 5.5: EFFECTS OF SCALING FOLIO ASSESSMENT AGAINST O-GRADE MARKS (BY SCHOOL)

School	*Folio Raw*		*Folio Scaled*		*Raw Folio match with O-grade*	*Scaled Folio match with O-grade*	*r Folio/ O-grade*
	MEAN	SD	MEAN	SD			
11 (149 pupils)	57·10	11·11	58·15	16·24	±1 range 48·4% ±2 ranges 71·8%	±1 range 42·8% ±2 ranges 68·3%	·68 (calculated for 143 pupils)
13 (262)	52·90	12·24	47·43	22·15	±1 range 37·6% ±2 ranges 58·9%	±1 range 42·6% ±2 ranges 59·7%	·71 (262)
14 (212)	54·24	9·79	53·56	19·35	±1 range 41·0% ±2 ranges 68·6%	±1 range 50·2% ±2 ranges 71·4%	·74 (196)
15 (163)	55·39	14·26	52·74	18·75	±1 range 45·2% ±2 ranges 71·4%	±1 range 45·1% ±2 ranges 67·7%	·81 (154)
21 (110)	60·90	10·98	58·08	15·72	±1 range 57·6% ±2 ranges 84·6%	±1 range 64·8% ±2 ranges 82·8%	·83 (93)
22 (236)	52·61	13·28	59·11	18·87	±1 range 41·2% ±2 ranges 65·1%	±1 range 54·7% ±2 ranges 76·1%	·79 (204)
23 (109)	52·14	14·41	51·74	20·04	±1 range 58·8% ±2 ranges 75·3%	±1 range 58·4% ±2 ranges 71·5%	·79 (81)
24 (73)	51·60	13·75	55·75	19·70	±1 range 54·8% ±2 ranges 75·3%	±1 range 46·5% ±2 ranges 67·0%	·84 (70)
32 (281)	55·90	11·94	52·39	18·32	±1 range 40·9% ±2 ranges 68·7%	±1 range 49·1% ±2 ranges 66·9%	·71 (261)
33 (59)	54·37	8·79	52·03	14·58	±1 range 44·0% ±2 ranges 72·8%	±1 range 44·1% ±2 ranges 62·8%	·39 (45)
35 (131)	57·48	10·98	61·38	18·62	±1 range 39·7% ±2 ranges 57·3%	±1 range 42·0% ±2 ranges 63·0%	·59 (112)

This table, containing information of several different kinds presents some problems of interpretation: no obvious pattern strikes the eye. It is possible, however, to note a number of points for consideration in trying to judge in what conditions scaling of Folio Assessment against standardised O-grade marks is likely to increase the number of pupils given the same or nearly the same ranges in both assessments.

1. The figures in the column headed 'Scaled Folio match with O-grade' indicate the amount of mismatch due to disagreements on ranking.

In schools obtaining Folio O-grade correlations of the normally acceptable order of ·70 or better, after scaling has removed the effects of the schools' severity or leniency and of their idiosyncratic distributions of marks, up to 40% of the pupils still obtained marks in one assessment which differed by two ranges (or one full band) or more from those obtained in the other.

2. The scaling process caused the match of ranges in some schools to be worse than before (see schools 11, 15, 23, 33; and also 21 and 24, where it was improved for ± 1 range but was worse for ± 2 ranges).

Distinct improvements in the match occurred in schools 13, 14, 22 and 35. Is it possible to identify any characteristics common to schools where the match was improved? The following sections seek to answer this question.

3. What is the effect on the match of ranges if the scaling process causes a large shift at the mean?

The largest mean differences occurred in schools 22 (6·5 marks), 13 (4·57 marks), 24 (4·05 marks) and 35 (3·9 marks). In three of these cases there was indeed an improvement in the match, and in school 22 it was a quite marked improvement. In school 24, however, the number of pupils obtaining the same or nearly the same range in Folio and O-grade was *smaller* after scaling. It may also be noted about those schools where the shift at the mean was no more than about 2·5 marks at most – schools 11, 14, 21, 23 and 33 – that all except school 14 suffered a drop after scaling in the number of pupils obtaining the same or nearly the same range in both assessments. It appears, then, that when scaling corrected a noticeable amount of severity or leniency at the mean it generally led also to an increase in the numbers of pupils given the same awards by Folio and O-grade. Certain other factors must, however, have been influential in school 24 and may also have contributed to the poorer range matching after scaling in those schools which had not required large adjustments to correct severity or leniency.

4. One of the other critical factors influencing the match of ranges awarded is the effect of scaling on the dispersion of marks, which can be judged by comparing the raw and scaled standard deviations of each school in Table 5.5.

The size of the adjustment made by scaling to the dispersion of marks seems to be of some significance to the match of ranges, especially if a large adjustment of dispersion is combined with a fairly large shift at the mean. In schools 13 and 35 there were shifts at the mean of some four or five marks and a much wider dispersion of marks after scaling than before, with improvements in range matching (though only a small improvement in school 13). The difference between raw and scaled standard deviations in these two schools is of the order of 8-10; it was only about 4-6 in all the schools which had poorer range matching after scaling. This means that the latter schools had spread out the pupils more effectively than the others in the raw Folio Assessment, and had, for instance, awarded more band As. Scaling will have moved some pupils within band A to other ranges, e.g., range 5 to range 3, range 4 to range 1, so increasing the count of pupils misplaced by one or more ranges. By the criterion of accurate assignment of ranges, therefore, schools using a wider spread of marks in Folio Assessment may *appear* to be less good judges of standard than others, if no attention is paid to the ability levels at which ranges in Folio and O-grade are matched.

A comparison between the effects of scaling at three ability levels in schools 14 and 15 will illustrate this point. Scaling made small differences to the means in each school, but a much greater difference to dispersion in school 14 than in school 15, so that in school 14 pass rates at bands A and 'F' were very different afterwards. Table 5.6 shows that in school 14 scaling brought dramatic increases in numbers of pupils obtaining bands A and 'F' in both Folio and O-grade, but not in school 15, where the teachers had already made quite a good job of assigning pupils to the extreme bands.

TABLE 5.6: SCHOOLS 14 AND 15: EFFECTS OF SCALING AT THREE ABILITY LEVELS

School	*Numbers obtaining Folio Bands before and after scaling*			*O-grade bands actually obtained by these pupils*					
		No.	*Pass Rate*	*A*	*B*	*C*	*D*	*E*	*'F'*
14	Folio 'A' Raw	= 11	= 5·1%	7	3	1	0	0	0
	Folio 'A' Scaled	= 43	= 19·8%	26	12	5	0	0	0
	Folio 'C' Raw	= 97	= 44·7%	12	21	28	18	11	7
	Folio 'C' Scaled	= 48	= 22·1%	4	16	13	1	4	2
	Folio 'F' Raw	= 3	= 1·4%	0	0	0	0	1	2
	Folio 'F' Scaled	= 24	= 11·1%	0	1	1	2	7	13
15	Folio 'A' Raw	= 32	= 19·6%	24	7	0	1	0	0
	Folio 'A' Scaled	= 34	= 20·1%	24	8	0	2	0	0
	Folio 'C' Raw	= 50	= 37·7%	1	12	17	13	3	4
	Folio 'C' Scaled	= 34	= 20·1%	2	9	10	11	1	1
	Folio 'F' Raw	= 15	= 9·2%	0	0	3	1	7	4
	Folio 'F' Scaled	= 11	= 6·7%	0	0	2	1	4	4

Table 5.6 also illustrates the typical effect of the scaling procedure on average pupils. In schools 14 and 15 and in all the others the number of pupils obtaining band C in both Folio and O-grade was smaller after scaling than before.

The main effect of scaling was to increase the match at the extreme bands and decrease it in the middle of the curve.

5. The effects of scaling on individual pupils in schools 14 and 15 do not depend solely on differences of mean and dispersion. The agreement on ranking between Folio and O-grade also has a significant influence. The respective correlation coefficients (school 14: ·74 for 196 pupils; school 15: ·81 for 154 pupils) are not significantly different. Both correlations would be accepted as satisfactory between two assessments of English, yet scaling had different effects on individual pupils in the two schools. It is clear that a high correlation between Folio and O-grade marks does not necessarily mean that scaling will be effective. Correlation coefficients may hide other important factors. The reader is referred to Appendix 8, which contains an examination of the results of the two schools showing the most favourable and unfavourable effects of scaling and which goes some way to defining the circumstances in which the scaling procedure applied is helpful. School 22 showed the greatest increase in match of ranges after scaling: school 24 suffered from more discrepancies after than before.

In brief, the study of these two schools shows that, though both have respectably high correlations between Folio and O-grade and between Folio and Criterion Test, mismatch in ranking between Folio and the other measures is distributed in different ways. In school 24 one class in particular is badly ranked by comparison with the Criterion Test and another of lower attainment is extremely well ranked and strongly influences the overall correlation; in addition, the former has been assessed leniently and the latter a little severely. The effect of scaling was to exaggerate the ranking disagreements in the poorly ranked class, and to cause some injustice to the pupils in the other class, because the scaling process, in trying to counterbalance the leniency shown to the better class, dealt with the less able one even more severely. (The details can be seen on pages 176-177.)

School 22, on the other hand, had been approximately equally severe at all ability levels and had also maintained a consistent degree of agreement/disagreement on ranking across the school. Scaling, therefore, had a very beneficial influence, adjusting the overall severity fairly and pushing the best pupils up to the top marks and the poorest down to the lowest.

The scaling procedure will cause some inaccurate awards to individual pupils unless the agreement in ranking is consistent across the school, i.e., unless there are no special 'class effects' in the School rank order.

III. What Sort of Reference Test ?

Since, in the SCRE project, Folio Assessment and O-grade were designed to test the same curriculum, the difficulties caused for the reference test method by variability of curricula did not arise in a noticeable way.

If schools took the opportunity of developing courses not closely related to the requirements of the external examination, the problem of the validity of a reference test for scaling internal assessments would be critical. It can be seen, however, from the correlation coefficients tabled in Appendix 13 that all the five separate elements of the Criterion Test measured achievements also measured by the Folio Assessment: a reference test containing, e.g., the two or three Criterion Test elements correlating most highly with the Criterion Test as a whole would probably cover most of the areas of written work English teachers would wish to engage in. (They would not, however, be suitable reference tests for, e.g., oral or drama work.) It can be shown in fact that combinations of any two Criterion Test elements in the SCRE project produce at worst a correlation of ·66 with the Folio Assessments (for a total of 1529 pupils). Combination of a writing task (either Criterion Test Paper IA or Criterion Test Paper III) with Criterion Test Paper II (traditional interpretation) or Criterion Test Paper IV (multiple choice interpretation) leads to correlations between the combined marks and the Folio Total of about ·70. Table 5.7 shows the correlations between particular elements in the Criterion Test and the Folio total mark and between combined elements of the Criterion Test and the Folio total. Any of the *combinations* of papers shown except that of Criterion Test Paper III (specified writing tasks) with Paper IV (multiple choice test) can in general be said to have measured enough of the same qualities as the Folio Assessment did to make their use as reference tests justifiable, and it is likely that the addition of a literature test to the Paper III plus Paper IV combination would bring its correlation with Folio total to over ·70. Two points should, however, be remembered.

1. These correlations are for a large number of pupils: some individual schools would achieve less satisfactory ones.

2. The calculation[1] of the correlation between two combined papers and the Folio Total mathematically determines 'weighting coefficients' or multipliers, to be applied to the marks in each paper in order to maximise the correlation of

---

[1] By formula $r_{1.23} = \sqrt{\dfrac{r^2_{1.2} + r^2_{1.3} - 2r_{1.2}r_{1.3}r_{2.3}}{1 - r^2_{2.3}}}$

See Q. McNemar (1962), page 175.

the combined papers with the Folio Total. These 'weighting coefficients' are given (approximately) in the column so headed, having been calculated by the method given by Q. McNemar (1962). In devising a reference test including two tasks like those of different Criterion Test papers, it would be necessary to give appropriate attention to the relative weightings of the two tasks in order to achieve the maximum correlation between the reference test and the internal assessment. The 'weighting coefficients' obtained here may give a lead in trying to determine what weight should be given to different test elements when combined with various others, though they would be absolutely appropriate only in the case where the correlations between internal assessment and the two separate reference test elements were identical to those obtained in the project for Folio against each relevant Criterion Test element.

TABLE 5.7: CORRELATIONS BETWEEN FOLIO AND CRITERION TEST ELEMENTS WHICH MIGHT FORM REFERENCE TESTS

Criterion Test elements	*Correlation with Folio Total*	*Approximate Weighting Coefficients* (β) *for combined papers*
Paper 1A – free composition	·61	
Paper III – specified writing tasks	·59	
Paper 1A and B – free composition and literature	·69	
Paper II – traditional interpretation	·66	
Paper IV – multiple choice interpretation	·58	
Papers 1A and II – free composition – plus traditional interpretation	·715	Paper 1A = 1·5 Paper II = 1
Papers III + II (specific writing tasks plus traditional interpretation)	·709	Paper III = 1·5 Paper II = 1
Papers 1A + IV – free composition plus multiple choice interpretation	·700	Paper I = 1·6 Paper IV = 1
Papers III and IV – specified writing tasks plus multiple choice interpretation	·683	Paper III = 1·2 Paper IV = 1
Papers 1A and B and II – free composition and literature plus traditional interpretation	·737	Paper 1A + B = 1·5 Paper II = 1
Papers 1A and B and IV – free composition and literature plus multiple choice interpretation	·737	Paper 1A + B = 1·8 Paper IV = 1

Note: Correlations are for 1529 pupils. Correlations for combined papers have been calculated by the formula for multiple correlations involving three variables.

If a reference test is used to scale internal assessment, it is obviously economical to reduce the test and the marking to the minimum necessary. Multiple-choice tests are attractive for this reason, provided that they are validly testing significant English skills. Consideration of correlations involving the marks for Criterion Test Paper IV are, therefore, of special interest, as they give some indication of the relationship between skills measured by that test and those measured by others.

First, how does the multiple choice test compare with the traditional interpretation test as a suitable reference test for Folio Assessment? Their respective correlations with Folio marks (for 1529 pupils) were ·58 and ·66 which are found to be significantly different when submitted to the 't-test' for the significance of the difference between two correlations with a criterion when the variables are also intercorrelated; i.e., the traditional interpretation paper in the Criterion Test measured a larger proportion than did the multiple-choice test of whatever the Folio Assessment measured. When the same two tests were correlated with O-grade marks, the traditional interpretation produced a coefficient of ·72 and the multiple-choice test one of ·63 – again a significant difference in favour of the traditional test.[1]

It will be seen from Table 5.7, however, that when either interpretation test is combined with a writing or writing and literature test, and appropriate weighting adjustments are made, the correlations between the combined writing and interpretation marks and the Folio Assessment are similar. Part of the explanation for this apparent contradiction is that there is a closer relationship between the writing task of, e.g., Paper IA and the traditional interpretation test ($r = \cdot 56$) than between writing (IA) and the multiple-choice test ($r = \cdot 45$). The traditional interpretation test had a greater overlap with the writing test, so that combining the latter with it raises the correlation with the Folio less than does combining writing with the multiple-choice test.

In sum, it would seem that a traditional interpretation test would be a better reference test than a multiple-choice one, if only the interpretation test were used. If, however, writing or writing and literature are combined with the interpretation test – which would be desirable – a wide enough range of English skills is covered whether the interpretation test is of the traditional type or in multiple-choice form, provided that suitable attention is given to adjusting the relative weightings of writing and interpretation according to the nature of the tasks set.

A reference test need not necessarily be applied to all pupils. A system in which a limited selection of common tasks from a suitable proportion of pupils' folios was actually marked externally by SCEEB could establish pass rates and the appropriate dispersion of marks for each school. The sample of pupils from each school would have to be a *substantial* one to ensure adequate representation of all levels of achievement.

Moderation of standards and spread of marks by a reference test would probably be satisfactory, but it would not solve the problem posed by schools (such as schools 13 and 35 in the project) in which the teachers produced a rank order markedly different from one which other assessors would draw up. If the SCRE Criterion Test rank order is taken as the criterion, the teachers' rank order in no school in the project could be said to be better than the O-grade one, while in schools 13 and 35 the O-grade ranking was clearly superior. This fact is apparently in conflict with the widely held view that teachers rank pupils best. They may do so within classes, but not necessarily on a whole-school basis.

---

[1] It should be noted that the traditional test required the pupils to write a summary; a fair evaluation of the multiple-choice questions would compare them with the traditional test without the summary.

It is likely that the best way of reducing the effects on comparability of lack of conformity of judgment about pupils' work is to combine internal assessment with external, so that some account is taken of both the teachers' and the examiners' judgments, and a counterbalance is applied to any extreme divergence on the part of the former from the consensus as to quality in English.

A system of combined internal and external assessment (the advantages of which are set out more fully in Chapter VII) would allow the possibility of a variation of the monitoring technique described above: work for external marking could be extracted from *every* pupil's folio, so that the extracts became, in effect, the external examination. The SCEEB mark for these extracts could be the reference mark to serve as a moderating touchstone for the rest of the pupils' work, and it could also be combined with the internal mark. The logistics and empirical effects of such a scheme would, however, require extensive investigation prior to implementation.

CHAPTER VI

MODERATION: SUMMARY OF FINDINGS

1. The Visiting Moderator scheme failed to identify one school which proved to be exceptionally severe in its assessments overall and two others in which the correlation between Folio marks and Criterion Test marks was shown to be significantly poorer than that for O-grade marks with the Criterion Test. It also failed in all eleven schools to adjust the numbers obtaining awards at each band to match the O-grade pass rates. This latter inadequacy could, however, be excused, since there were clear indications that the numbers falling into each band in the O-grade English examination are determined to a significant degree, not by the judgments of the markers as to the standards achieved, but by the statistical standardising procedure, which spreads out the usually very bunched distribution of O-grade marks.

The implication of these failures of the Visiting Moderator scheme is that the SCRE project has not established that such a scheme would guarantee comparability of standards of marking across schools within the normally accepted limits of tolerance, which are severity or leniency not exceeding one SCE range per candidate. If comparability of marking standards and of pass rates at each band is the priority, a common reference test is needed.

2. The O-grade examination provided such a reference test for the project. *Scaling Folio marks against O-grade marks did effectively correct overall severity or leniency in the schools and affected the schools' dispersions of marks so that pass rates at each band in Folio were comparable with those in O-grade.*

3. The effect of scaling on internal awards to individual pupils depends partly on the degree of correlation between Folio and O-grade. *When the correlation was ·70 or better there remained still up to 40% of a school's candidates who received internal marks (after scaling) differing by the width of one full band or more from their O-grade marks. Combining internal and external assessment may be the best way of dealing with disagreement on ranking.*

4. It was found, further, that even in schools with fairly high correlation between Folio and O-grade scaling against O-grade could lead to inaccurate awards for individual pupils unless the quality of the school assessment had been consistent in all classes. *In-service training, including Trial Marking exercises, would be necessary to ensure that teachers were aware of the need for assessments to be standardised within each school. Adherence by the teachers to instructions to produce a school rank order rather than to assess each class separately would be critical. It will be seen in the following chapter that a* common *interpretation test is probably also requisite.*

5. Central to the problem is the SCEEB standardising procedure, which, in the case of O-grade English, makes a significant difference to the numbers of pupils awarded high and low bands. *The typical bunching of O-grade English marks both in the examination and in internal assessment, may be due to markers' and teachers' unwillingness to award high and low marks – the researchers certainly did note a tendency to unrealistic idealism when teachers were asked in the preliminary stages of the project what they expected A pass pupils to be able to achieve – or it may reflect the true distribution of English skills in the sixteen-year-old population as far as it is possible to distinguish one pupil's command of*

them from others'. If the latter is the case, distinctions between pupils obtaining contiguous ranges and even between those obtaining bands B, C and D in either Folio or O-grade are much less clear-cut than the standardised ranges and bands suggest, and markers and teachers should be taught to think of discriminating, not in terms of ranges, or, with even stronger reason, of continuous marks out of 30, 50 or 100, but in terms of bands.

6. *It might have been better to scale Folio marks against* raw *O-grade marks. This would have effected any necessary correction of overall severity or leniency and it would have standardised dispersions of marks, but by making smaller adjustments than were made using the method employed in the project. Disagreements on ranking would not then have been so grossly exaggerated by the stretching out of three-mark categories to five-mark ones.*

7. *The results of comparisons between the SCRE Criterion Test elements and both Folio Assessment and the O-grade examination suggest that an external reference test should contain at least writing and interpretation elements.*

8. *If the interpretation test is a multiple-choice one, a literature element is desirable as well as composition, but is not essential. The combination of all three types of task, if appropriately weighted, covers a range of skills comparable to that measured by Folio Assessment in the project; the combination of multiple-choice interpretation with free composition alone, though not so satisfactory, also correlated adequately with Folio Assessment.*

CHAPTER VII

TWO FACTORS AFFECTING THE QUALITY OF ASSESSMENTS

I. Inter-Marker Inconsistency

The knowledge that the marking of English compositions suffers from inter-marker inconsistency is now relatively ancient. The findings of Hartog and Rhodes in *The Marks of Examiners* (1936) and of the International Institute Examinations Enquiry in *The Marking of English Essays* (Hartog *et al*, 1941) have in fact had beneficial influences on examination marking in English which are still effective. It was the latter report which encouraged a concern for judgments in the first instance about the *sense* of the writing, this being assessed in terms of the writer's success in attaining his object. It also advised against the practice of arriving at a final mark for a composition by summing marks for separate elements such as spelling, vocabulary, structure . . . , a warning which is still repeated in the SCEEB instructions to markers. Despite advising means of improving the assessment of writing, the distinguished authors of *The Marking of English Essays* were so aware of the unreliability of examination essays as a means of testing the ability to write that they recommended a trial period during which the writing task would be removed altogether from the School Certificate Examination (with the safeguard for its continuance in schools that Boards of Education might have powers to reduce grants if composition were not taught!). They were pessimistic, having failed to obtain any assurance from their evidence that examiners all meant even approximately the same thing by the term 'literacy', having noted great divergencies between the marking of the same scripts by different examiners and, in a re-marking exercise, having found that the examiners disagreed not only with one another but also with their own previous judgments. French sources are quoted to reinforce their own findings that different experienced examiners are quite capable of awarding the same candidates marks which differ by more than 25% of the possible total.

That such problems were still current was suggested by the arguments for multiple-marking of compositions made out in *Multiple Marking of English Compositions* (Britten, Martin and Rosen, Schools Council, 1966).

The SCRE researchers, therefore, approached the question of inter-marker unreliability relating to the assessments under investigation not in order to establish anything new about marking in English, but with a view to obtaining some information about the degree of success with which modern procedures for standardising the judgments of markers reduce the effect of unreliability. Dr Ballard, one of the authors of the 1941 report, wrote in *The New Examiner* (Ballard, 1929): 'The examiner's blunders are as carefully hidden from the public gaze as the doctor's blunders under the tombstones.' It was not necessary to agree with the implication in that statement of deliberate cover-up to feel that an appraisal of current marker-inconsistency in SCE O-grade English would be worthwhile, if it served only to remind all concerned of the continuing prevalence of the problems which caused concern to Hartog and his colleagues forty years ago.

A study was accordingly made of variations between assessments of the same script by different markers of the Criterion Test. As the markers of this test were all experienced and satisfactory O-grade markers, who were performing for the project a task very similar to both O-grade marking and Folio marking, conclusions may be drawn from this study, with due caution, about likely effects of marker variation on both school and SCE assessments.

The usual procedure in a study of comparability between marks would be to arrange for all the markers involved to mark a common set of scripts. The SCRE Criterion Test procedures did not allow this but did permit an approximation to a multiple-marking arrangement.

Those Criterion Test Papers which were to be double-marked (i.e., I, III and V, the writing and literature assignments) from each school were divided into eight alphabetical groups and, as far as possible, each First Marker received one bundle from each school. The papers were afterwards distributed to the Second Markers in such a way that one Second Marker received work marked by each of the previous markers. This enabled the researchers to compare the marks of each marker with those of eight other markers who had each marked a portion of his scripts. In the following discussion the usual assumption has been made that groups of pupils allocated to each marker contained the full range of O-grade candidates.

The pattern of differences between the marks awarded by two markers to the same pupils showed up not only consistent differences in severity or leniency but also other ways in which marking behaviour varied; differences in dispersion of marks and disagreement on ranking, which may be called 'inter-marker inconsistency'. Tables 6.1 and 6.2 below show marker characteristics for Criterion Test Paper I (Composition and Literature), Paper III (Factual and Persuasive Writing) and Paper V (Short Story). The columns headed 'Mean', 'Standard Deviation' and 'Inter-marker Inconsistency' show measures for each marker which relate to the three ways in which comparability between markers is affected. Differences at the mean may have been caused by consistent severity or leniency. Usually, however, one would expect inter-marker comparability to suffer also from the differences in the dispersions of marks which are indicated by the varying standard deviations. These differences of means and of dispersions of marks are susceptible to statistical adjustment. There are, in addition to them, differences between markers which are due to genuine differences of opinion as to the quality of a script or to factors such as markers' carelessness, loss of concentration, or haphazardness in applying standards themselves. The figure under 'Inter-marker Inconsistency' is the standard deviation of the differences between each marker's score for each of his batch of papers and the score for the same paper given by the other marker of the eight who acted as second marker for that paper. This measure of inter-marker inconsistency is affected by the varying dispersions of marks among the markers, but is not directly dependent on them. Its usefulness is in showing that all markers differ from their colleagues to approximately the same degree. One marker only, 34 (see Table 6.2) was a 'rogue', showing considerably greater disagreement with the others than any other marker.

TABLE 6.1: MARKER CHARACTERISTICS ON CRITERION TEST PAPER I (MARKED OUT OF 60)

Marker	*n*	*Mean*	*S.D.*	*Inter-marker Inconsistency*
1	264	31·1	7·85	6·10
2	258	32·5	7·47	5·01
3 (1st Marking)	266	32·7	9·85	5·81
4	262	33·7	7·30	5.25
5	184	30·0	7·82	4·89
6 (1st Marking)	275	30·1	6·20	4·35
7	279	28·4	8·99	5·73
8	301	32·0	9·62	6·12
3 (2nd Marking)	225	29·4	7·63	5·10
6 (2nd Marking)	182	27·9	5·99	5·13
23	258	30·6	6·76	5·48
25	236	28·2	6·92	4·97
26	265	30·8	7·47	5·47
27	252	33·4	8·60	5·82
28	259	30·7	8·73	6·50
29	253	32·0	8·46	4·82
30	223	28·0	7·60	4·59

Notes: (1) Marker 30 attended the second markers' meeting but his allocation of scripts included some '1st marking' work left over because of the illness of marker 5.

(2) Though there was no difference in the standard set by 1st and 2nd markers on the sample scripts at the markers' meetings, the overall mean for 1st Marking was 31·3, and for 2nd Marking 30·3. Means on sample scripts were respectively 30·7 and 31·1.

(3) Markers 3 and 6 did a double stint, and their characteristics appear to have changed between 1st and 2nd Marking.

TABLE 6.2: MARKER CHARACTERISTICS ON CRITERION TEST PAPERS III AND V. (MARKED OUT OF 20 EACH)

	PAPER III				PAPER V		
Marker	*n*	*Mean*	*S.D.*	*Inter-marker Inconsistency*	*Mean*	*S.D.*	*Inter-marker Inconsistency*
17	262	10·37	3·23	2·82	10·57	3·56	2·59
18	264	10·64	3·08	2·67	10·18	3·41	2·71
19	260	9·72	3·05	2·46	9·24	4·06	2·93
20	252	10·53	2·82	2·67	8·95	3·38	2·80
21	240	10·66	2·47	2·68	9·32	3·48	2·81
22	253	9·83	3·01	2·71	9·76	3·62	2·96
23	263	11·24	2·27	2·91	10·88	3·39	2·66
24	266	9·78	3·11	2·89	7·04	3·57	2·72
1	253	10·07	3·73	2·73	9·68	3·81	2·88
21	225	10·40	2·54	2·30	9·13	3·22	2·57
31	265	11·00	3·13	2·05	9·01	3·62	2·23
32	255	7·58	2·91	2·38	7·90	2·77	2·53
33	210	8·68	2·95	2·21	8·29	4·18	2·86
34	281	9·30	4·39	3·51	7·06	4·85	4·04
35	267	9·73	2·94	2·45	8·82	2·95	2·48
36	250	9·42	3·39	2·71	8·52	3·62	2·40

Notes: (1) Marker 1 also did First Marking of Paper I.

(2) Paper III, overall means: 1st Marking = 10·3; 2nd Marking = 9·5.
Paper III means at markers' meeting: 1st Marking = 11.6; 2nd Marking = 11.5.
Paper V, overall means: 1st Marking = 9·5; 2nd Marking = 8·5.
Paper V, means at markers' meetings: 1st Marking = 10·5; 2nd Marking = 10·6.

Tables 6.1 and 6.2 indicate disagreements among markers in statistical terms only: it is of interest to note what they mean in real terms for pupils under assessment, and this may be expressed by showing the discrepancies between marks awarded to the same script by First and Second Markers. Tables 6.3 and 6.4 show the figures.

TABLE 6.3: CRITERION TEST, PAPER 1

DIFFERENCE BETWEEN FIRST AND SECOND MARKS BEFORE MARKER-STANDARDISATION

(a) Paper 1A – Composition – out of 30 marks.

Maximum difference = 12 marks (5 cases)

No. of marks different	*% of available marks**	*No. of pupils*	*% of all completed papers*
0 to 2½	less than 10%	1112	52·0%
3 to 5½	10 – 19%	744	34·8%
6 to 8½	20 – 29%	227	10·6%
9 to 12	30% or more	56	2·6%
		2139	100%

* Note: It was rare for an essay to be given less than 10 marks, so that in effect only 20 marks were used for this paper. Percentages given are, however, based on the possible total of 30 marks.

(b) Paper 1B – Literature – out of 30 marks.

Extreme differences = 13½ (1 case), and 13 (3 cases)

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 2½	less than 10%	1217	56·8%
3 to 5½	10 – 19%	647	30·3%
6 to 8½	20 – 29%	221	10·3%
9 to 13½	30% or more	56	2·6%
		2141	100%

(c) Paper 1 – A and B – out of 60 marks.

Maximum difference = 25 marks

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 5½	less than 10%	1460	68·3%
6 to 11½	10 – 19%	586	27·4%
12 to 17½	20 – 29%	84	4·0%
18 to 25	30% or more	6	0·3%
		2136	100%

TABLE 6.4: CRITERION TEST, PAPERS III AND V

DIFFERENCE BETWEEN FIRST AND SECOND MARKS BEFORE MARKER-STANDARDISATION

(a) Paper III – Factual and Persuasive Writing – out of 20 marks.

Extreme differences = 11½ marks (1 case), 9 (4 cases)

No. of Marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 1½	less than 10%	952	46·5%
2 to 3½	10 – 19%	673	32·8%
4 to 5½	20 – 29%	339	16·5%
6 to 11½	30% or more	87	4·2%
		2051	100%

(b) Paper V – Response to short story – out of 20 marks.

Maximum difference = 11 (3 cases)

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 1½	less than 10%	963	47·1%
2 to 3½	10 – 19%	642	31·5%
4 to 5½	20 – 29%	305	15·0%
6 to 11	30% or more	130	6·4%
		2040	100%

One point about Table 6.3 serves to convey the significance of all these figures. *As a result of various combinations of severity, leniency, different spreads of marks and inter-marker inconsistency, 31·7% of 2136 pupils, if marked by only one of their two markers, would have obtained on Criterion Test Paper I marks differing by 6 – 25 out of 60 from what the other marker would have awarded.* Criterion Test Paper I was similar in form and marking instructions to O-grade Paper I (though worth 60 rather than 50 marks because of its extra Literature question). The 16 markers constituted as experienced and satisfactory a group as could be found. The standard deviation of raw O-grade marks being usually about 12·5 – 13, six or seven raw marks are the equivalent of two ranges or one band. The implication for the reliability of the marking of Writing and Literature, both in O-grade and in any internal assessment, is clear. *It is quite possible for about 25% of candidates to obtain an O-grade result which would have differed by one full band if their Paper I had been marked by another marker. In the case of a further 3% or 4% the discrepancy would be two full bands. It is, of course, very creditable to the O-grade examining team that some 70% of over 70,000 candidates are reliably assessed in a subject notorious for marker-subjectivity. Nevertheless, if the SCRE Criterion Test study is indeed indicative of O-grade marking behaviour, in absolute terms, some 17,000 – 20,000 pupils probably have their O-grade Composition and Literature answers assessed with a significant degree of unreliability before marker-standardisation.*

Considerable time and effort is spent in 'marker-standardisation', which is a process by which the Principal Examiner and his colleagues seek to check that every marker is working consistently to the standards they set. Their method has two stages. First, scrutiny of a small sample of each marker's scripts, resulting in a judgment as to overall severity/leniency; a plus or minus numerical factor is suggested at this stage to correct the tendency if the standardisers think that there is a consistent trend and that the marker has not treated any particular ability level specially. Later, another numerical factor is provided by the computer, when each marker's mean score has been compared with the national average mark. A decision is then taken as to the factor to be applied to the marker's awards, consideration being given to his apparent consistency and his spread of marks. There is, however, no standardisation of the dispersion of marks. If the application of a numerical factor would benefit some pupils and harm others, no standardisation of that marker actually occurs, unless he is so inconsistent as to warrant the remarking of all his scripts. In O-grade English, Paper I, it would be unusual for a marker to receive a factor greater than ± 3 raw marks.

There is also a scrutiny of scripts 1 – 2 marks below the C/D borderline, to try to ensure that no one 'fails' as a result of severe marking.

The effect of this moderation process on the results tabulated above is an important issue. If consistent severity or leniency were a major cause of discrepancies between two sets of marks for the same scripts, the effect of SCEEB marker-standardisation could be expected to be significant. If varying spreads of marks were largely to blame, standardisation of dispersion, as well as of mean, would be needed to improve matters. No standardisation procedure other than re-marking by the Principal Examiner's team would be effective if the main cause of the problem were inter-marker inconsistency.

When comparisons were made between first and second markings for small groups of pupils it was found that the pattern of discrepancies was seldom such that adjustment of the mean alone would noticeably improve agreement, and cases in which standardisation of dispersion of marks led to *much* closer agreement were also infrequent. All too often the main characteristic of the graphs drawn was disagreement on ranking, even when the markers' inconsistency as measured by the figures given in Table 6.1 did not mark them out specially from their colleagues. It therefore seemed unlikely that the process of marker-standardisation would make great difference to the numbers of pupils awarded alarmingly discrepant marks by the two markers of their scripts.

That this was in fact the case can be seen from Tables 6.5 and 6.6 which show the improvement due to marker-standardisation, and which may be compared with Tables 6.3 and 6.4. (Although the pupils involved here are a subset of those in the earlier tables, the figures are still comparable.)

TABLE 6.5: CRITERION TEST PAPER I
DIFFERENCE BETWEEN FIRST AND SECOND MARKS AFTER MARKER-STANDARDISATION

(N.B. Slightly different population from Table 6.3)

Paper 1A – Composition – out of 30 marks.

Extreme differences = 14 (1 case), 13 (1 case), 12 (2 cases)

No. of Marks different	*% of available marks*	*No. of pupils*	*% of completed papers*
0 to 2½	less than 10%	924	49·0%
3 to 5½	10 – 19%	747	39·5%
6 to 8½	20 – 29%	186	10·0%
9 to 14	30% or more	28	1·5%
		1885	100%

Paper 1B – Literature – out of 30 marks.

Extreme differences = 12 (1 case), 11 and 11½ (5 cases)

No. of marks different	*% of available marks*	*No. of pupils*	*% of completed papers*
0 to 2½	less than 10%	993	52·6%
3 to 5½	10 – 19%	693	36·8%
6 to 8½	20 – 29%	164	8·7%
9 to 12	30% or more	35	1·9%
		1885	100%

Paper 1A and B – out of 60 marks.

Extreme differences = 21 (1 case), 17 (1 case), 16 and 16½ (2 cases), 15 and 15½ (6 cases)

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 5½	less than 10%	1320	70·0%
6 to 11½	10 – 19%	496	26·3%
12 to 17½	20 – 29%	68	3·6%
18 to 21	30% or more	1	0·1%
		1885	100%

TABLE 6.6: CRITERION TEST, PAPERS III AND V

DIFFERENCE BETWEEN FIRST AND SECOND MARKS FOR EACH PUPIL AFTER MARKER-STANDARDISATION

Paper III

Extreme differences = 8 and 8½ (8 cases), 7 and 7½ (15 cases)

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 1½	less than 10%	705	37·4%
2 to 3½	10 – 19%	795	42·4%
4 to 5½	20 – 29%	312	16·6%
6 to 8½	30% or more	73	3·9%
		1885	100%

Paper V

No. of marks different	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 to 1½	less than 10%	691	36·7%
2 to 3½	10 – 19%	767	40·7%
4 to 5½	20 – 29%	328	17·4%
6 to 8½	30% or more	99	5·2%
		1885	100%

The reduction effected by marker-standardisation in the number of pupils awarded marks discrepant by 10% or more of the available marks by two individual markers on Paper IA and B is of the order of 1% – 2%, and on Papers III and V the effect was to increase the numbers of pupils unreliably assessed. (It may be that the extremely aberrant performance of one marker, 34, had exceptional influence.) The average discrepancy between First and Second markers per pupil on Paper IA, Composition, was 3 to 3·5 marks (out of 30) both before and after marker-standardisation.

These findings have important implications for O-grade English marker-standardisation. The SCRE marker-standardisation, which equalised means and dispersions of marks, if applied to a population as large as the O-grade one – about 70 000 pupils – would have led to a reduction in blatant unreliability of mark for 1% – 2%, or about 1000 pupils. About 30%, or some 20 000 pupils, would still have been left with a mark differing by six or more out of sixty from the one another marker would have given. SCEEB marker-standardisation probably has even smaller beneficial effect than this, because it does not affect dispersions of marks. The question arises whether the annual hard work on marker-standardisation by the Principal Examiner and his team has a significant enough effect to justify it.

It may be felt that the labour of marker-standardisation is justified if even a few pupils thereby achieve a fairer award. It does not seem likely that modification of the procedures would lead to noticeably more effective marker-standardisation, since the bulk of discrepancies between markers appear to be in that category of inconsistency, haphazardness, or genuine difference of opinion about the quality of individual scripts which precludes the regulation of markers' awards by statistical adjustment. The procedure in current use could continue to pick out very obviously aberrant markers – though it had apparently failed to note over several years the distinct lack of comparability between the judgments of the marker numbered 34 in the SCRE project and those of most other markers.

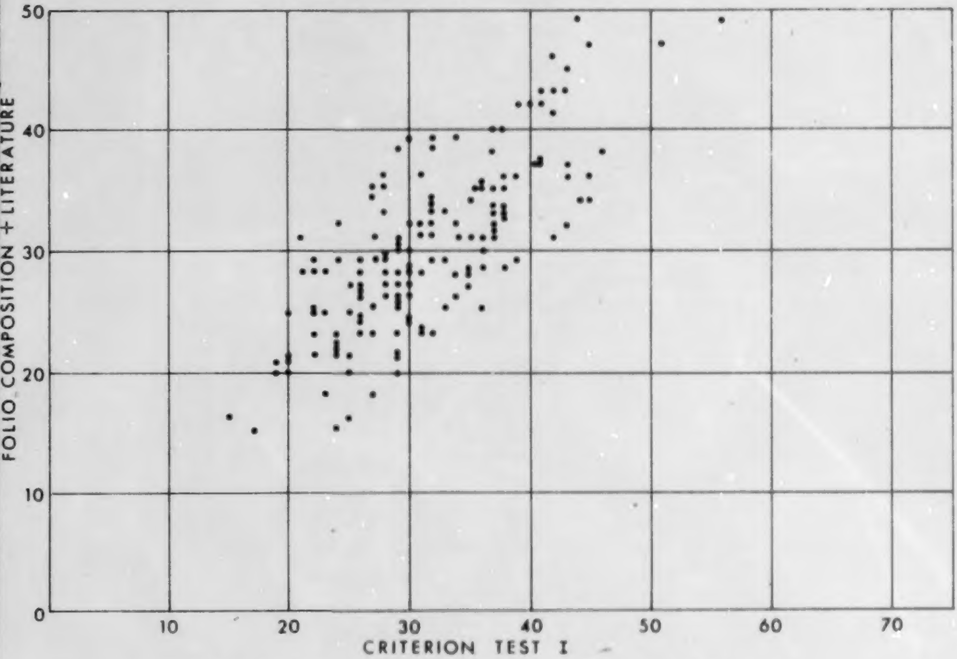
Are there alternatives? If the external examination of Writing and Literature is sacrosanct, there is a strong case for double impression marking as described by Wood and Quinn (1976) as a means of reducing the effects of inter-marker inconsistency. Also attractive, however, is the combination by simple addition of internal and external assessments, provided that both are measuring some, at least, of the same aspects of 'English'. Such an arrangement has the advantages of simultaneously helping to combat marker-inconsistency and taking account of different but equally valid responses to pupils' writing, by increasing the number of judgments about that writing which contribute to the assessment[1].

It can be shown that combining several assessments of writing and literature does lead to closer agreement between marks given by teachers and examiners on the one hand and a criterion measure on the other. Figure 6.7 contains a series of graphs in which one assessment is plotted against one other (Fig 6.7(a) and (b)), two assessments added together are plotted against one other (Fig. 6.7(c) and (d)), and two similarly combined assessments are plotted against two others, also added (Fig. 6.7(e)). The narrowing 'cloud' of plotted points shows that combining the Folio and O-grade assessments increases their agreement with the Criterion Test assessments of writing.

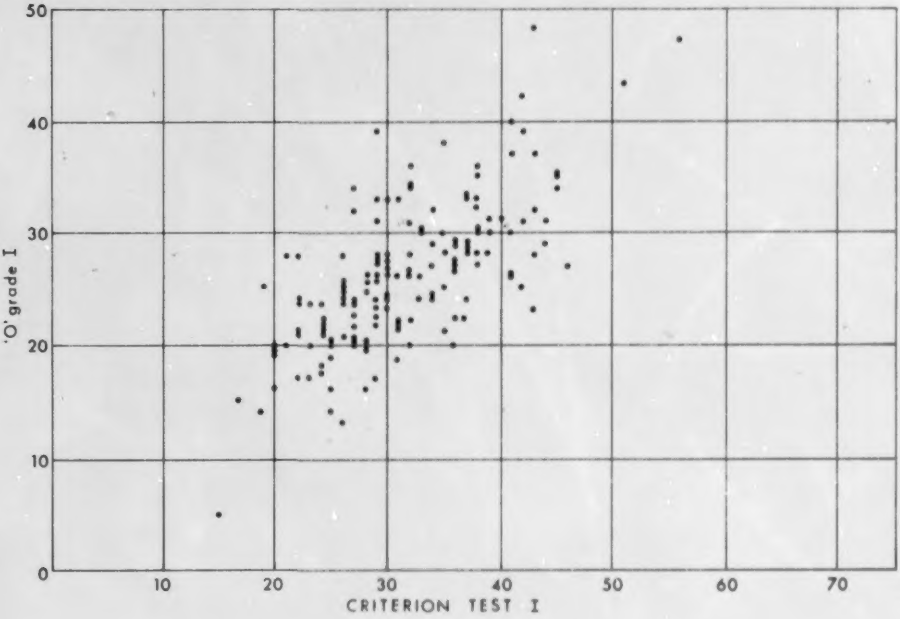
---

[1] See Britton, Martin and Rosen, Schools Council (1966).

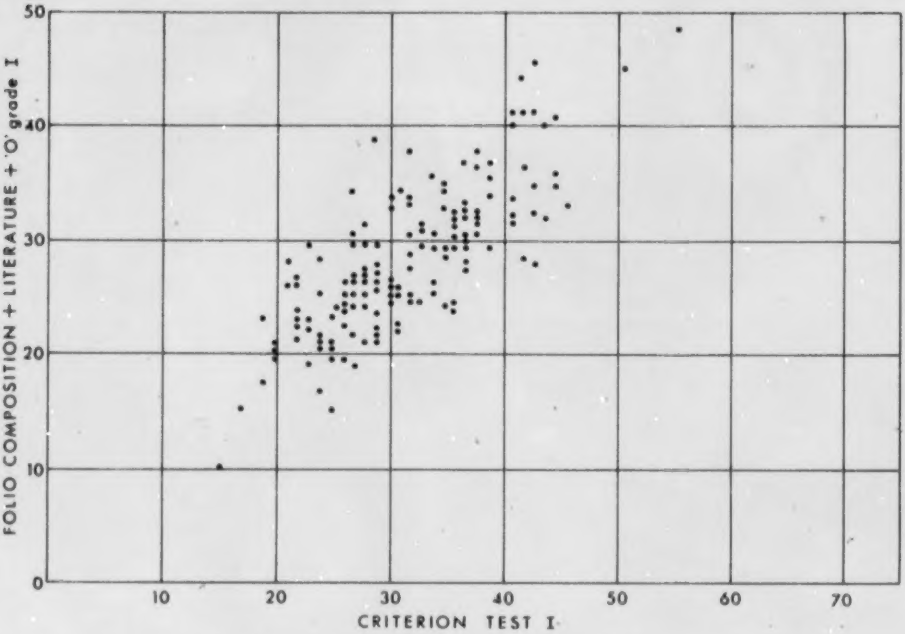
FIGURE 6.7 SCHOOL 15: COMBINED ASSESSMENTS
a) Composition and Literature in Folio against one Criterion Test paper



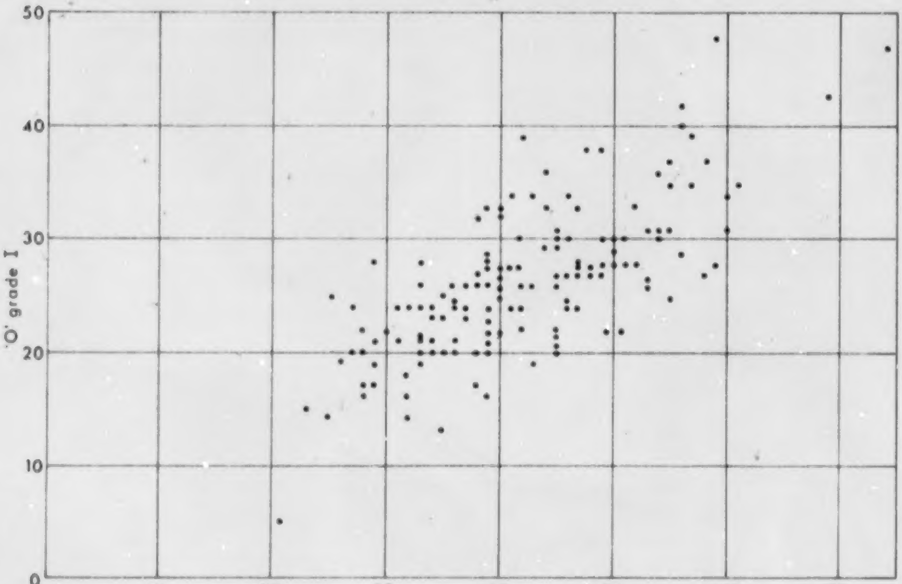
b) One O-grade paper against one Criterion Test paper



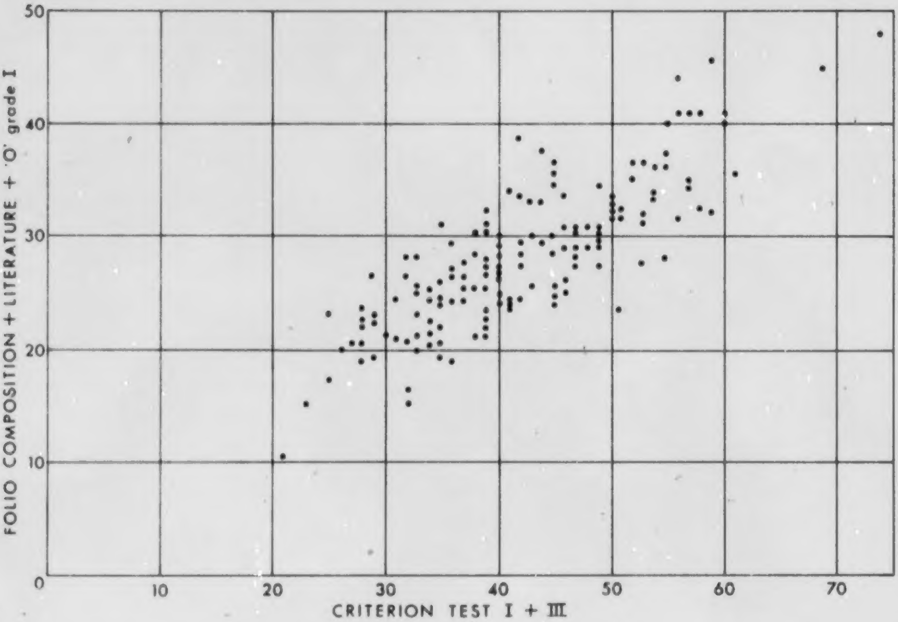
c) Folio and O-grade I against one Criterion Test paper



d) O-grade Paper I against two Criterion Test papers



e) Folio and O-grade I against two Criterion Test papers



The effect was also tried of combining Folio and O-grade totals (i.e., Writing, Literature and Interpretation) for comparison with the Criterion Test totals. Does the combination of Folio and O-grade correlate better than either alone with the Criterion Test? Table 6.8 shows the various Pearson correlation coefficients for each school. Two means of combining the Folio and O-grade individual correlations were employed. The first, designated in Table 6.8 as 'Folio plus O-grade against Criterion' is that of averaging the two correlations[1]. The other, called 'Folio × O-grade against Criterion' is calculated following a formula which adjusts the relative contributions of Folio and O-grade so that the correlation obtained between Criterion Test and the combined Folio and O-grade is the best possible from any linearly additive combination of the latter two[2].

[1]

$$r_{1,(2+3)} = \frac{\sigma_2 r_{1,2} + \sigma_3 r_{1,3}}{\sqrt{\sigma_2^2 + 2r_{2,3}\sigma_2\sigma_3 + \sigma_3^2}}$$

[2]

$$r_{1,23} = \sqrt{\frac{r_{1,2}^2 + r_{1,3}^2 - 2r_{1,2}r_{1,3}r_{2,3}}{1 - r_{2,3}^2}}$$

TABLE 6.8: COMBINED FOLIO + O-GRADE v CRITERION TEST

School	*n*	*Folio against Criterion Test*	*O-grade against Criterion Test*	*Folio against O-grade*	*Folio × O-grade against Criterion Test*	*Folio plus O-grade against Criterion Test*
11	143	·74	·69	·68	·78	·75
13	163	·68	·85	·61	·87	·86
14	218	·83	·81	·74	·88	·88
15	154	·84	·84	·81	·88	·88
21	93	·86	·82	·83	·88	·88
22	104	·84	·85	·79	·89	·89
23	81	·84	·87	·79	·90	·90
24	70	·89	·88	·84	·92	·92
32	268	·83	·81	·71	·89	·89
33	45	·55	·71	·39	·77	·76
35	112	·63	·84	·59	·85	·82

Table 6.8 shows that the best possible combination of Folio and O-grade was barely better than simple averaging of the two. In all cases except school 35 the average of Folio and O-grade correlated slightly better with the Criterion Test than either Folio or O-grade alone, though the differences were not large. While the advantages of combining Folio and O-grade are not very great as compared with O-grade alone, if it is desirable to have an internal assessment for educational reasons, it is heartening to note that even in the two schools where the Folio assessment correlated significantly less well than O-grade against the Criterion Test – schools 13 and 35 – the combined assessments were distinctly better than the relatively poor Folio assessment alone and not significantly different from the O-grade as compared with the Criterion Test. It would seem that the effect of unsatisfactory characteristics influencing the Folio marks in these two schools was noticeably reduced by the simple averaging of them with O-grade marks.

Equally weighted internal and external assessments would, on this evidence, probably be satisfactory.

A Note on the Nature of the Writing Task and Reliability of Marking

A reconsideration of inter-marker inconsistency in the Criterion Test will find indications that there was probably less disagreement on ranking among the markers of Paper III, in which specific tasks were set with defined purposes and audiences, than in the free composition of Paper IA. Table 6.9 compares the two sets of figures.

TABLE 6.9: MARKER-INCONSISTENCY ON PAPER IA AS COMPARED WITH PAPER III

PAPER IA (marked out of 30)				PAPER III (marked out of 20)			
No. of marks different Marker 1 – Marker 2	*% of available marks*	*No. of pupils*	*% of all completed papers*	*No. of marks different Marker 1 – Marker 2*	*% of available marks*	*No. of pupils*	*% of all completed papers*
0 – 2½	less than 10%	1112	52%	0 – 1½	less than 10%	952	46·5%
3 – 5½	10 – 19%	744	34·8%	2 – 3½	10 – 19%	673	32·8%
6 – 8½	20 – 29%	227	10·6%	4 – 5½	20 – 29%	339	16·5%
9 – 12	30% or more	56	2·6%	6 – 11½	30% or more	87	4·2%
		2139	100%			2051	100%

A strict comparison between the marking of the two tests would take account of the real, as opposed to the nominal, weightings of each. Without calculating real weightings exactly, it can be noted that the full scale of 20 marks was used by Paper III markers, whereas it was rare for a script to receive less than 10/30 on Paper I. The two tests actually had, therefore, roughly comparable mark scales, and, in absolute terms, five or six marks on one had approximately the same value as five or six marks on the other. A legitimate comparison can then be made between the percentages of pupils with discrepancies between Marker 1 and Marker 2 of six or more marks. On Paper I the figure was 13·2%, on Paper III, only 4·2%.

The arguments against one option composition tests are that they may be advantageous or disadvantageous to some pupils and that they would probably have the effect of limiting the kinds of writing undertaken in schools. In the present state of knowledge, one does not know whether the gains or losses to individual pupils due to lack of options would be greater or less than those caused now by marker-inconsistency resulting from too many options: it would be an interesting piece of research to find out. The effect on curriculum might be lessened if internal assessment of writing and literature were given significant weight in the overall mark for English and pupils were obliged to produce for assessment in school writing of various kinds. The curriculum could then actually lead assessment, as the Dunning Committee (SED, 1977) and many others have wished, and teachers could use assessment to establish whether particular skills they have sought to help pupils develop are in fact exhibited. Lip-service is frequently paid to this model of assessment but it is probably non-existent in Scotland for S4 English. If English teachers really believe in it, it is their responsibility to declare: 'These are the kinds of writing and literature we wish to teach: let them be given significant weight in assessment.' The external examination would then be freed from its present impossible role of trying to achieve reliable discrimination among pupils while at the same time seeking to encourage, or at least allow, in schools an infinitely wide variety of courses, texts and emphases in English teaching.

So long as significant weight were given to internal assessment of several types of writing, it would be possible to set a single clearly defined task, without choice, in the external paper to facilitate greater reliability of marking. Variation from year to year of the nature of the external test would avoid the danger of schools devoting inordinate amounts of time to preparing pupils for one kind of writing only. If, however, the external test is also to be used to moderate the standard and dispersion of marks of the internal assessment, care would be necessary to ensure that it did test a fair proportion of the writing skills also tested by the teachers. Table 5.7 (page 50) shows that tasks like those in Criterion Test Papers I and III combined with a traditional interpretation test form a reasonably satisfactory reference test for Folio Assessment: further research would be required to establish the suitability or unsuitability of any other kinds of specific task, such as speech writing, before they were employed.

II. The Influence of the Task and Choice of Tasks

(a) Writing and Literature

Evidence that a wide choice of questions or assignments contributes to unreliability in examination marks generally has been gathered by several researchers, most recently by Willmott and Hall (1975) in ''O' Level Examined: the Effect of Question Choice'. There is, however, some difference of opinion as to the effects of question choice in English examinations. Willmott and Hall refer to J. M. Stalnaker's essay on 'The essay type of examination' in Lindquist (1951), *Educational Measurement*, in which he argues that allowing a choice of essays implies the assumption that all the topics would equally well facilitate the measurement of some general ability to write. If this dubious assumption is correct, no optional questions are necessary. If it is not so, and the successful pupil can in fact write well only on the one topic he selects, 'the significance of the possession of this ability is so difficult to interpret that one may question the use of the results'. (Willmott and Hall, 1975, page 8.) On the other hand, Wiseman and Wrigley (1958) found that the real differences which occurred between the means of children selecting different essay topics in an English examination were largely due to differences of ability in the children, and that topic choice in English composition was unlikely to introduce substantial error into marking. Wiseman and Wrigley point out that abolition of choice might or might not affect reliability adversely. More recently, Adams and Pearce (1975) quote H. Rosen's PhD thesis for London University as having established that a pupil's grade is determined as much by his choice of composition topic as by any other factor. Stevens (1970) shows that grades awarded to some 'A' level English Literature candidates depended to a significant degree on their choice of questions. Research does not appear to have settled the matter.

It is suggested in the other section of this chapter that the vagaries of markers might be more easily kept in check if all pupils tackled the same specified writing task, or, preferably, several such tasks. One of the advantages of Folio Assessment is that it can test performance in several kinds of writing and so reduce whatever problems are caused by allowing choice of style and topic in a single examination.

At the same time as Folio Assessment reduces choice-of-writing-topic problems, it unfortunately seems to introduce a factor detrimental to comparability of assessment which is not present in an external examination: a variety of interpretation tasks for different pupils under the same assessment.

The discussion of bunching and marker effects in section I of this chapter has concentrated on Writing and Literature. It is appropriate to give some attention to the credentials of Interpretation as a testing tool before commenting on difficulties of its use in Folio Assessment. (The term 'interpretation' is here used as synonymous with 'comprehension'.)

(*b*) *Interpretation Testing*

Given passages of suitable difficulty and validly constructed questions, an Interpretation test ought to discriminate more finely and more reliably than composition: marks are acquired, not on the basis of a holistic impression, but step by step as individual questions are correct, and it is possible to limit the effect of the markers' subjectivity by a predetermined marking scheme or one agreed among the markers. There is undoubtedly some marker inconsistency in the marking of O-grade Paper II and similar tests set in schools, but probably considerably less than in the marking of Composition. The researchers could not carry out for Criterion Test Paper II a cross-marking study like that made for the double-marked Papers, but can show that there was variation of both standard and dispersion of marks among the eight very experienced markers of this test. Examples of different distributions of marks are given in Appendix 9, and the means and standard deviations for all eight markers can be seen in Table 6.10.

70

TABLE 6.10: MARKER CHARACTERISTICS ON CRITERION TEST PAPER II

Marker	*Mean*	*S.D.*	*No. of scripts*
9	20·1	7·0	283
10	18·4	6·8	273
11	18·3	6·9	347
12	18·4	6·6	290
13	20·0	7·3	287
14	19·5	7·3	291
15	21·7	9·3	102
16	20·2	8·1	260
Overall	19·3	7·3	2133

Note: The figures for markers 15 and 16 are after the application of a – 1 marker-standardisation factor.

A feature of the consideration of the marking of Criterion Test Paper II was the fact that some markers produced an overall mean comparable to that of their colleagues by distinct severity in marking the summary and distinct leniency on the rest of the Paper, or vice-versa. Multiple-marking of a random sample of O-grade or Criterion Test Paper II scripts would be a valuable exercise which would help in the formation of rational judgments about the possible use of multiple-choice tests at O-grade.

Whatever the reliability of Interpretation tests, their main disadvantage is that they are not easy to construct validly. Problems arise in choosing passages of appropriate style, content and difficulty for the standard of the examination and for the range of pupils taking it. It can also be argued that Interpretation tests, as traditionally set, are unrealistic reading tasks which are not valid tests of pupils' general reading ability, because nobody ever reads anything in the way the tests require, except in examinations. Writing unambiguous valid questions at the right level of difficulty is another tricky job. The case that schools should assess reading in some way other than the traditional interpretation test could certainly be put. The difficulties of ensuring comparability of interpretation tests set in different schools and different classes had been expected: the researchers believe that they have in fact found some statistical evidence, discussed below, which shows the detrimental effects for accurate ranking of non-comparability of interpretation tasks used as the basis for a common assessment.

FIGURE 6.11 SCHOOL 24:
Lenient Assessment of Class 2

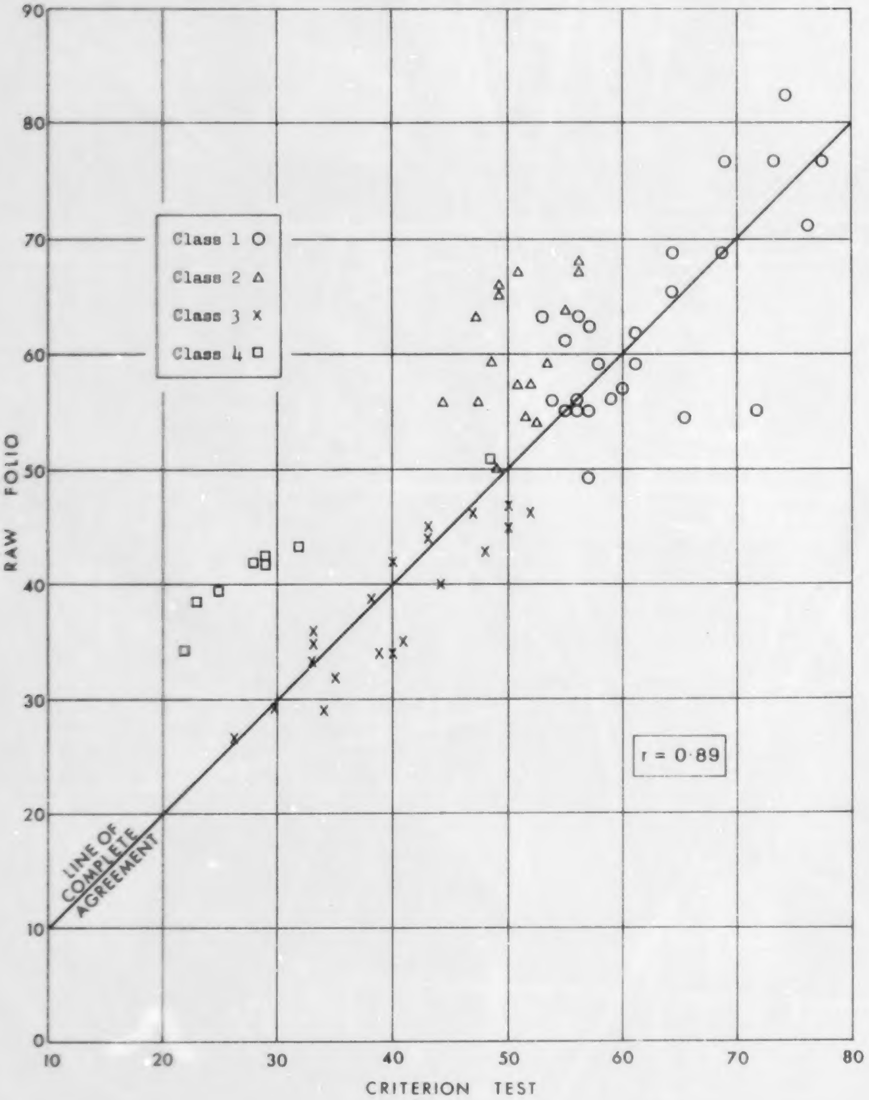
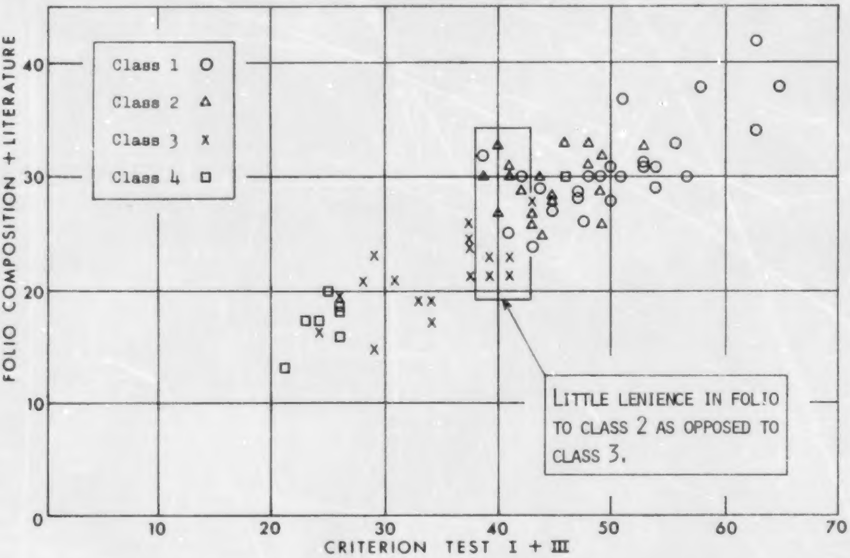
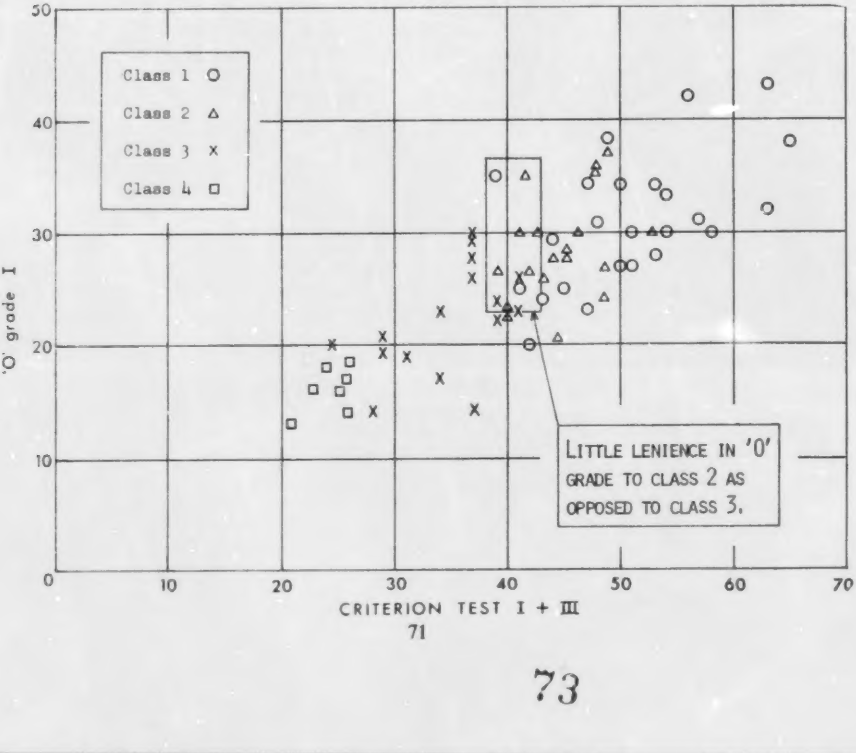


FIGURE 6.12a SCHOOL 24:
Folio Composition and Literature against
Criterion Test Composition and Literature



LITTLE LENIENCE IN FOLIO TO CLASS 2 AS OPPOSED TO CLASS 3.

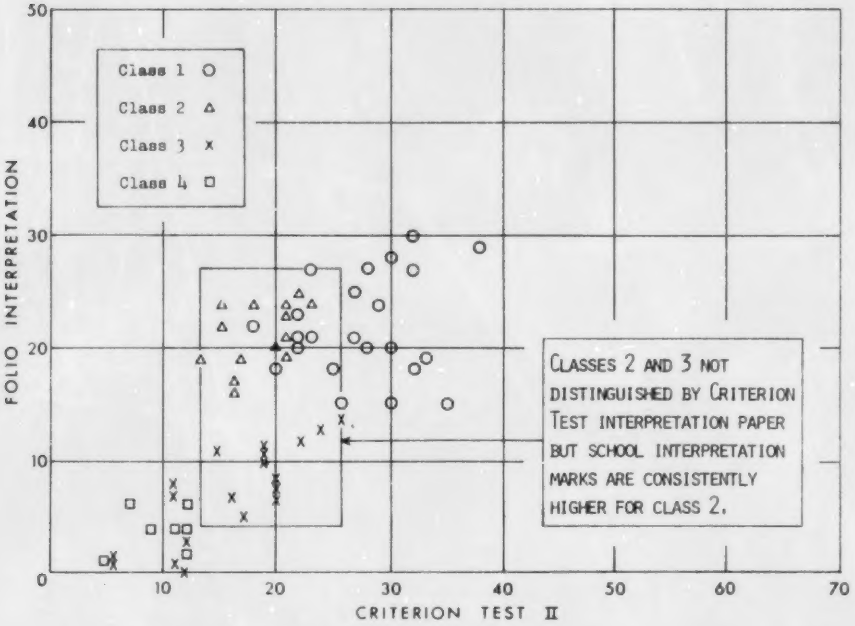
FIGURE 6.12b SCHOOL 24:
O-grade Composition and Literature against
Criterion Test Composition and Literature



LITTLE LENIENCE IN 'O' GRADE TO CLASS 2 AS OPPOSED TO CLASS 3.

FIGURE 6.12c SCHOOL 24:

Folio Interpretation against
Criterion Test Interpretation



There were 'class effects' in a number of the 'whole-school' assessment schools, possibly due to a tendency to give higher marks to all pupils in top classes and low ones to all pupils in bottom classes, or to failure to adhere to the instruction to mark on a team basis. Another cause could also be identified. In order to save teachers' time, the instructions for marking Folios did suggest that the marker, assessing the folio work of pupils he had not taught, could make a speedy judgment about the standard of interpretation marking previously carried out by the class teacher on work in the folio. If he was satisfied that it was in fact of O-grade standard, he could accept as valid any interpretation marks given by the class teacher, without himself having to re-mark every question. In effect, then, the teachers were required to mark only writing and literature as a team: they could, and probably did, accept the class teachers' interpretation marks. Any other arrangement would in fact have been impractical. 'Class effects' may then have been caused by varying standards of interpretation tasks and marking between classes. The clearest example of the effects of different interpretation work in different classes is provided by school 24. Figure 6.11 shows that class 2 in that school was more leniently assessed on Folio than other classes. Figures 6.12(a) to (c) prove that the leniency was mainly in the assessment of Folio Interpretation. Figure 6.12(a) plots Folio Composition and Literature marks against those for the Writing and Literature elements in the Criterion Test. Only a few pupils in class 2 overlapping on Criterion Test marks with some of class 3 were given higher Folio Composition and Literature scores than the class 3 pupils. The same applies to the assessment of writing and literature in the O-grade (see Figure 6.12(b)). Figure 6.12(c) however, shows that on Criterion Test Paper II, Interpretation, classes 2 and 3 were not clearly distinguished from one another, whereas class 2 obtained consistently higher marks for Folio Interpretation than class 3.

The combined difficulties of lack of time to standardise team marking of all possible interpretation tests set in various classes and of setting passages and questions of comparable difficulty in different classes and different schools lead to the tentative conclusion that, if interpretation tests are to be retained as a means of assessing reading ability on a national basis or on a school basis, it would be advisable for all pupils to take common tests, nationally, or at least within each school.

The conclusion is tentative because the researchers are not convinced that a single national interpretation test is, as is commonly believed, a valid and reliable test of reading generally. It is as hard for the setter of O-grade Paper II to produce a thoroughly valid test as it is for the teachers, given the difficulties listed earlier in this section – and the O-grade has to cater for 70,000 candidates. Marker unreliability affecting Paper II is as yet unexplored.

It has been suggested in the preceding chapter (page 51) that the common external interpretation test should preferably not be a multiple-choice test if it is the *only* reference test employed. If, however, it is combined with writing and literature tests, the interpretation test, on the evidence of the SCRE Criterion Test results, may equally well be traditional or multiple-choice: in either case the combination of writing, literature and interpretation tasks covers a range of reading and writing skills wide enough to form a satisfactory reference test for Folio Assessment, provided that appropriate attention is paid to the relative weightings of each element.

The SCRE data could probably yield more information than has so far been obtained about the quality of current methods of testing interpretation. In the meantime a system of common interpretation tests seems preferable (as a means of discriminatory testing) to one allowing each teacher to find or invent his own. Such a system would not, however, preclude schools from assessing reading performance, by means of traditional interpretation work or by other means such as the tasks set in Criterion Test Paper V, and combinations of marks might be appropriate; e.g., an external examination mark might be combined with an internal impression mark for various kinds of interpretation work.

CHAPTER VIII

SUMMARY OF STATISTICAL EVALUATION

1. Seven out of eight schools engaged in 'whole-school assessment' produced a rank order of pupils not significantly different from that of the O-grade examination in assessing achievements also measured by the SCRE Criterion Test; in 'class assessment' schools, one out of three did so.
2. Whole-school assessment is desirable, rather than class assessment. A common standard of marking within the school is desirable, as well as awareness of standards elsewhere acquired through, e.g., Trial Marking Exercises.
3. Organisational and management difficulties, especially lack of time for assessment, probably adversely affected the quality of internal assessment in some schools. Similar problems also prevented three of the experimental schools from co-operating in the project as agreed, and caused a fourth to provide only partial data.
4. In the assessment of writing and responses to literature the following factors probably contributed to unreliability in both Folio and O-grade.
    - (a) Inter-marker inconsistency, causing some 25% of pupils to obtain marks a full band or more different from those they would have received from a different marker. Marker-standardisation had little influence on this inconsistency.
    - (b) Lack of perceived distinguishing qualities in pupils' scripts, so that a very bunched distribution of marks resulted. (O-grade results as publicly reported do, however, give an illusion of clearer discrimination between pupils because of the SCEEB scaling procedure.) English teachers and markers should learn to think of discrimination, especially among the large mass of 'average' pupils, in terms of the five or six bands rather than in terms of continuous marks out of 50 or 100.
    - (c) Variety of choice of tasks/stimuli.
5. In the assessment of interpretation (or comprehension), the following points emerged from the researchers' experience.
    - (a) There is a need for further study of the validity and reliability of this traditional method of testing reading ability.
    - (b) In the meantime, if comparability is desired, common interpretation tests are preferable to a variety of tasks set by different teachers in different classes or schools.
6. Schools' pass rates at each band differed from O-grade pass rates after standardisation of O-grade marks. The schools awarded too many Cs and too few As and 'F's. This mismatch with O-grade standards was correctable, *en gros*, by the application to each school's Folio marks of the mean and standard deviation of its marks in the examination.
7. This scaling process significantly increased the match between Folio and O-grade awards in particular schools only in the circumstances where the school's ranking had been comparable with that of the O-grade examination *at all ability levels*, but there were discrepancies between Folio mean and O-grade mean and between Folio dispersion of marks and O-grade dispersion. Disagreements on ranking were exaggerated by the scaling process.

    Alternative approaches should be investigated to see if the exaggeration of ranking disagreements could be reduced while mean and dispersion are appropriately standardised.

8. An external reference test against which to-scale internal assessments like the Folio Assessment in this project should contain at least writing and interpretation elements. If appropriately weighted the interpretation element could consist of a multiple-choice test.

9. Provided that they are all validly assessing some common aspects of 'English', combinations of assessments, especially of Writing and Literature, can improve discrimination and reliability.

CHAPTER IX

PRACTICAL PROBLEMS

I. Problems in Obtaining Accurate Data

Failures by the schools, for whatever reasons, to provide all the required data, or to send it in the desired form, may indicate the kinds of communication problems and recording errors likely to arise in a real internal assessment system. They are important also because they may bear on the degree of correlation among the three assessments made in each school.

(*a*) *Inability to provide data*

The effect of pressure of work on teachers, which is discussed in the following section, led to the failure of two schools to provide folio assessments at all, to one providing assessments for a proportion of pupils only, and to a fourth school being unable to administer the Criterion Test.

(*b*) *Inaccuracy*

Some inaccurate reporting of scores and misinterpretation of instructions occurred in the other schools.

The schools had been asked to report three marks (composition, literature and interpretation) and assign a 'range' (1-14) based on the total of these, following the pattern of the scaled O-grade marks, i.e.:

range 1=90-100 2=85-89 3=80-84 4=75-79 5=70-74
6=65-69 7=60-64 8=55-59 9=50-54 10=45-49
11=40-44 12=35-39 13=30-34 14=0-29.

Four variables were therefore required (three marks, one range). Some teachers sent 5 variables: 3 marks, a total and a range.

Among some 2000 pupils, 227 discrepancies occurred between the total mark (whether recorded by the teachers or not) and the range assigned (e.g., a total of 54 being translated as Range 8 instead of 9). Only seven of these proved to be copying errors by SCRE staff.

So about 11% of the folio assessments from 10 schools were in some way unsatisfactorily reported to SCRE.

Some other errors arose because some classes in two of the Strathclyde schools had recorded the three components of the folio as ranges rather than marks. This probably occurred because of a changed instruction to the schools: they were asked to record their Autumn assessments as fourteen ranges, but for the later assessments three percentage marks and one (total) range were demanded, because it had become clear that the statistics would require a wider range of marks than 1-14.

With hindsight one can say it would have been better not to ask for ranges, but simply for the three component marks and the total out of 100. It would then have been desirable to check the totals immediately and ask the schools to clarify those marks in which there appeared to be an error.

(*c*) *Deliberate 'errors'*

When all these various reasons had been taken into account, substantial discrepancies between the total folio mark and the assigned range still remained for a considerable number of pupils in particular schools or classes. Unless some teachers were excessively careless, these were probably deliberate and, indeed, in some cases were explicitly indicated as such.

The cases of known deliberate discrepancy between Folio Total and Folio Range occurred in schools 31 and 35 and are specially interesting. They suggest that some teachers regarded their pupils' 'ability' in English as something other than what the work in their folders testified to. In some cases it was explicit that the teacher was making a prediction of performance in the O-grade in assigning ranges. The temptation to do this was stronger if pupils had failed to complete the minimum number of assignments for the folio assessment: hence some 'estimate marks'. A third factor mentioned in discussion and written documents by teachers at school 35 was the feeling they had that some of their pupils had under-performed in certain folio assignments, namely the 'close reading' exercises recommended as a possible 'different' sort of O-grade work: the teachers thought that unfamiliarity with this kind of work had caused some pupils to score marks lower than their 'ability' warranted.

Schools 31 and 35 may have been the only ones where such subjective judgments were made, or they may have been more honest than the others. There is some indication that the instruction to produce folio marks on a *school* basis was not strictly implemented in all the schools which received it.

It would seem that in any 'real' internal assessment scheme there should be a very clear instruction as to what exactly teachers are being asked to assess, this preferably being pupils' actual achievements in school rather than putative performance in an external examination. The continuing presence and influence of an external examination may well, however, make such an instruction difficult for teachers to adhere to.

II. Problems for Teachers

(*a*) *Ordinary nuisances*

The project must have been influenced by those ordinary annoyances of daily school life, well known to anyone who has ever taught a class, which drain away time and energy in disciplinary, organisational, and administrative matters. In some of the project schools the enthusiasm of both teachers and pupils for internal assessment was dampened by unfortunate timetabling (e.g. all S4 English classes last period in the afternoon, except one, which was last period in the morning!), unavailability of suitable folders, or vandalism of folders kept in insecure cupboards in S4's distant annex location. A few teachers seemed disheartened by absenteeism and lack of motivation in the 'less able' classes they were teaching and felt that the extra administrative and organisational work imposed by the project was merely that, and did not benefit the unwilling pupils.

(*b*) *Teachers' attitudes*

A second influence may have been the degree of willingness with which individual teachers became involved in the project in the first place. Some welcomed the idea of internal assessment and worked hard to produce material for the project. Others, drawn in with their colleagues or their Principal Teacher, were satisfied with the existing system and, envisaging nothing more than extra work resulting from the project, were, in a few cases, hostile to it from the start or, in others, indifferent to its success. A negative attitude on the part of the teachers was not widespread, but it was certainly to be found occasionally, and lack of interest, if not hostility, surely contributed to the failure of one whole school to provide the promised internal assessments.

(*c*) *Confusion of purposes in the project*

A number of difficulties arose because of the particular ways in which the project was conducted and from imperfect communication. Some were caused by the double intention of the researchers on the one hand to investigate the comparability of school assessments with the O-grade and the Criterion Test, and, on the other, to encourage more detailed course planning, perhaps involving some break with O-grade tradition, as the essential first step in developing an internal assessment system. The late arrival in the schools of the 'Guidance Booklet', with its advice on both the standardisation of assessments and the design of them, meant that some teachers felt that they had not had time to consider properly the requirements of the project. Some misunderstanding arose, too, over the place in the project of the 'close reading' which was suggested as a type of work not normally covered in O-grade classes and of which some exemplars were provided in the booklet. The intention of the researchers had been that marks for such work, if carried out, would be recorded under 'Literature' or, perhaps, 'Interpretation', but several teachers in one school complained that they had devoted considerable time to preparing suitable material only to find that the Folio Assessment did not take 'close reading' into account.

(Some other failures in communication relating to the recording of folio assessments are referred to in the preceding section on problems in obtaining accurate data.)

It is the view of the research team that confusion was created in some teachers' minds, not so much by the fact that the project was trying to standardise assessments and encourage course-planning at the same time, as by the fact that it failed to convince these teachers first that course-planning is part of assessment design and that it should precede choice of assessment technique. Despite the double intention, the emphasis in the project fell on the need to standardise judgments about pupils' performance rather than on design of an assessment procedure. This was, of course, partly because standardisation is necessary, but also because the researchers, who were obviously not able to insist on certain courses being followed, were obliged to settle for an assessment model already in operation – the O-grade one – because that model matched, and, indeed, determined the sort of work the pupils were going to do in the year. Although several teachers did modify their S4 teaching slightly, by excluding past papers or including more 'close reading', no school took up suggestions that folios might contain significant amounts of work arising from courses which were not directly derived from the format of the O-grade examination, e.g., a drama-based or role-playing-based course, or a course on the press. This fact is recorded without imputation of laziness or hidebound conservatism: the pupils were, in fact, taking the O-grade examination and their time and their teachers' time were limited.

It is arguable that assessment of pupils' performance in courses which varied greatly from school to school could not be standardised, as the marks might be measures of quite different achievements. If this were so, the criticism that the project was trying to do two separate and confusing things at the same time would be justified. The experience of CSE Boards suggests, however, that standardisable assessments of English skills exhibited in different types of course can be made. This project can make no contribution of evidence for or against that proposition.

(*d*) *Lack of time*

By far the most significant practical problem was pressure of time, both long and short term.

There was a general feeling, most intense in the Grampian region, that the preparatory planning time was too short. Initial contact was made with most of the schools in late February, 1976, for an August start to the experiment; the Grampian schools were approached in mid-March. The papers describing the proposed scheme were circulated in April and departmental meetings to discuss them were held with the Research Officer during that month. Agreement was reached then that the 14 schools would engage in the project as set out in Chapter II, and in June, when the first Trial Marking meetings were held, there was further discussion with the teachers about the procedures. Several teachers felt these arrangements were hurried and would have preferred a full year's notice. It should, however, be noted that the shortage of planning time may have been more acute for some teachers than for others because of breakdown in communication within their schools.

Whether the internal assessment scheme could have been better prepared by the schools in a year than in three to four months is debatable. A year's foreknowledge would not necessarily mean a year's preparation, because teachers' time is very easily filled up by matters of immediate concern. It seemed reasonable to the research team to suppose that one term in advance is as much as a department would normally allow for planning such a change, but perhaps a few months more, at least, would have helped teachers to feel more at ease when the internal assessment scheme started.

Whatever the amount of notice allowed, there is no doubt that the teachers were very hard pressed to fulfil the demands of the project and carry out their normal teaching duties simultaneously. Most of them did succeed in both, some spending many out-of-school hours in preparation of new teaching material and in assessing pupils' work. It is clear that enthusiasm and good use of the available time can make a success of an innovation. It would, however, be a distortion of the truth to report otherwise than that English teachers are carrying a very heavy burden of class preparation, correction and assessment, so that any additional work is really more than they can take on without detriment to some aspect of their teaching. There is no consensus of opinion, let alone research evidence, as to the time needed to plan English classes and assess written work, but there must be some doubt whether English teachers can *now* perform for all pupils all the functions they would regard as part of their job: they would then still be over-committed even if folio assessment tasks replaced some they already carried out, such as the marking of 'prelim' or end-of-term examinations. It is possible to obtain a rough impression of the demands on teachers' time from the following set of figures, which indicates approximately the amount of time per week English teachers in those of the project schools which provided this information can spend assessing the written work of each of their pupils.

In the course of normal English classwork there is, properly, a great deal of informal assessment and marking going on, and, also properly, pupils will produce some written work which does not need or is not meant for marking. Even when these points are remembered, the figures indicate the immense potential volume of written work an English teacher may receive each week for assessment outside of class contact time.

'Average assessment time per week' was calculated as the contractual 32½[1] hours, less class contact time, less two hours for class preparation. (Two hours is a very ungenerous allowance for this purpose, even if the teachers have managed to use some post-examinations free time the previous June to plan the main lines of their work for the current session.) Many teachers, of course, work at home for longer than the five hours per week which the 32½ hour week

---

[1] 27½ hours in school, 5 hours at home.

TABLE 7.1: TIME AVAILABLE FOR ASSESSMENT OF WRITTEN WORK

School	*No. of pupils*	*No. of teachers*	*Pupils per teacher per week*	*Average assessment time per week*	*Assessment time per pupil per week*
14	1700	11	155	7 hours = 420 mins (plus approx. 1½ hrs for travel between buildings)	2½-3 mins
21	1050	8·5	124	6·7 hours = 400 mins	3 mins
22	1700	13	131	6·3 hours = 380 mins	3 mins
23	800	7	114	7·6 hours = 455 mins	4 mins
31	1560	8·5	183	6·3 hours = 378 mins	2 mins
32	1700	14	121	7·7 hours = 460 mins	3½-4 mins
34	1150	8	144	7·7 hours = 460 mins	3 mins
35	1300	10·5	124	6·3 hours = 378 mins	3 mins
These 8 schools	10960	80·5	136	approx. 7 hours = 420 mins	3 mins

requires, and some exceptional teachers and departments can perform organisational miracles to use time more efficiently. Table 7.1 seems to suggest, nevertheless, a genuine need for considerably more free time for English teachers, if it is an important part of their work to assess written work either for diagnosis of strengths and weaknesses or for discrimination.

An alternative or additional ameliorating step would be to reduce class sizes further and so also the volume of written work per teacher.

While it is not the business of this project to make recommendations about pupil-teacher ratios, it is legitimate, in the context of discussion of the practical problems relating to internal assessment, to observe that in order to allow a teacher to spend *five minutes* per week assessing each pupil's output, within contractual working hours and with the average free time available in the eight schools in Table 7.1, the pupils taught by each teacher per week would have to number about 84 rather than about 136. This would reduce all classes to about 18-21 pupils and would require the employment of nearly 60 additional teachers in these eight schools.

There is at present a natural tendency for teachers to concentrate their marking effort on certificate classes. The implications for 'certification for all' of English teachers' shortage of time are clear.

As far as the project itself is concerned, it seems inevitable that, without staffing *changes, a reduction in the minimum amount of work to be assessed should be recommended for any future internal scheme, though enough should be included to ensure that a range of different types of work is covered.*

Despite the problems described above, it is not the researchers' opinion that internal assessment is impossible because of pressure of work on teachers. Some of the project schools, and many in England, have shown that an internal assessment programme can work very well. The important point about the shortage of time is that it applies whether or not assessment for certification is internal. If the English teacher's almost impossibly heavy workload cannot be reduced, it may well be possible to *change* some of it from direct preparation for and practice testing for the external examination to assessment of internally designed courses. The project schools were abnormally heavily burdened by having to do both at once.

CHAPTER X

TEACHERS' VIEWS, 1977

I. The Questionnaire

The 1975 survey of the views of over 500 English teachers produced the following division of answers to the question whether they would welcome suitably moderated internal assessment instead of the O-grade English examination.

TABLE 8.1: DESIRABILITY OF INTERNAL ASSESSMENT (1975)

Yes	*No*	*No Response*
51%	42%	7%

100% = 523 English teachers.

This survey was made in 104 schools, one in four of those presenting candidates in the O-grade English examination. It was not possible to know precisely the total number of English teachers in these schools, but it was reasonable to guess that the figure of 523 represents 60% – 70% of the total possible sample.

At the end of session 1976/77 the teachers who had participated in the internal assessment scheme were asked to fill in a questionnaire and the researchers visited each school to collect the responses and receive any word-of-mouth comments. Seventy-three of the 114 teachers involved completed at least some parts of the questionnaire; though they form a very different sample from the 1975 group, their opinions are of special interest because of their experience. Tables 8.2 and 8.3 show their responses to two sections of the questionnaire which listed possible advantages of course-work assessment generally, and of the 'rules' of the SCRE experiment in particular.

TABLE 8.2: ADVANTAGES OF COURSE WORK ASSESSMENT: TEACHERS' VIEWS (1977)

Advantages	*Occurred*	*Did not occur in project, but a potential advantage*	*Occurred, but not regarded as an advantage*	*Did not occur and unlikely*	*No response*
Wider range of work	14 (19·2%)	26 (35·6%)	1 (1·4%)	27 (37·0%)	5 (6·8%)
Work better suited to pupils	17 (23·3%)	17 (23·3%)	2 (2·8%)	24 (32·9%)	5 (6·8%)
Pupils better motivated	33 (45·2%)	13 (16·4%)	0 (0%)	20 (27·1%)	8 (11·0%)
Work of better quality	18 (24·7%)	15 (20·5%)	0 (0%)	30 (41·0%)	10 (13·7%)
Pupils more accurately assessed	24 (32·9%)	22 (30·1%)	0 (0%)	26 (35·6%)	1 (1·4%)
No time lost for 'prelim' examination	9 (12·3%)	5 (6·8%)	7 (9·6%)	42 (57·5%)	10 (13·7%)
Better course planning	10 (13·7%)	28 (38·3%)	0 (0%)	24 (32·9%)	11 (15·0%)
Useful teaching material produced	34 (46·6%)	13 (17·8%)	5 (6·8%)	12 (16·4%)	9 (12·3%)
Teacher more aware of standards and techniques of assessment	38 (52·1%)	15 (20·5%)	1 (1·4%)	12 (16·4%)	7 (9·6%)

100% = 73 teachers.

TABLE 8.3: ADVANTAGES OF SCRE PROJECT: TEACHERS' VIEWS (1977)

Project Element	*Very advantageous*	*Acceptable*	*Small advantage only*	*Definitely not advantageous*	*Undecided*	*No response*
Definition of aims, as principle	21(28·8%)	28(38·4%)	1(1·4%)	2(2·8%)	5(6·8%)	16(21·9%)
Actual aims agreed	14(19·1%)	29(39·7%)	1(1·4%)	1(1·4%)	7(9·6%)	21(28·8%)
Specification of minimum work, as principle	22(30·1%)	23(31·5%)	3(4·1%)	1(1·4%)	9(12·3%)	15(20·5%)
Actual minimum agreed	1(1·4%)	24(32·9%)	1(1·4%)	11(15·0%)	15(20·5%)	21(28·8%)
Takes account of *improved* work	19(26·0%)	18(24·7%)	6(8·2%)	8(11·0%)	7(9·6%)	15(20·5%)
Takes account of *best* work	19(26·0%)	25(34·2%)	4(5·5%)	4(5·5%)	8(11·0%)	13(17·8%)
'Trial marking' exercises	18(24·5%)	29(39·7%)	13(17·8%)	4(5·5%)	9(12·5%)	0(0%)
Moderator's visit to discuss levels of difficulty of work	26(35·6%)	17(23·3%)	6(8·2%)	2(2·7%)	17(23·3%)	5(6·8%)
Moderator's visit to discuss standard of marking	25(34·2%)	22(30·1%)	4(5·5%)	3(4·1%)	10(13·7%)	9(12·3%)

100% = 73 teachers.

Descriptions of the aims and other elements of the project referred to in Table 8.3 can be found in Chapter II. The list in Table 8.2 consists of advantages mentioned by teachers when the research team had visited the schools in the course of the year. In discussion during these visits it was noted that some teachers saw likely advantages of internal assessment which were not, however, manifesting themselves in the experimental scheme because of its 'teething troubles' and because of the constraints of time caused by having to prepare for the O-grade examination as well. It was for this reason that a category of 'potential advantages' was included in the questionnaire.

It is clear from the tables that, when they are considering the educational advantages for the pupils and their own professional interests, the weight of opinion among these teachers is favourable to internal assessment and to the form of it adopted by the SCRE project. In Table 8.2, responses in the 'Occurred' and 'Potential' columns outweigh those in the other two columns in every case, except 'No time lost for "prelim" examination', and, in Table 8.3 all the listed elements of the project received more support than criticism. It is noticeable in Table 8.3, however, that a much larger proportion of teachers made no comment on some particular aspects of the project, especially 'Actual aims agreed' and 'Actual minimum agreed'. They may have felt, as the researchers did, that it was difficult to judge the value of some of the requirements of the project in an experiment lasting only two terms and carried out under the pressure of the need to prepare pupils for the external examination.

The impression that the majority of the questionnaire respondents are favourably disposed to folio assessment is reinforced by consideration of their reactions to most of the items which invited judgments about possible drawbacks. Table 8.4 shows that most teachers felt that pupils had not suffered or, in a smoothly operating scheme, would not suffer from the disadvantages, which, again, had been suggested to the research team by teachers during the course of the experiment. The one clear disadvantage was the tendency to concentrate on S4 work to the detriment of other classes; and the fairly large numbers who found or envisaged difficulty in keeping folders intact and in order are noteworthy. These are both management problems rather than educational issues.

TABLE 8.4: DISADVANTAGES OF FOLIO ASSESSMENT FOR PUPILS: TEACHER'S VIEWS (1977)

Disadvantages	*Occurred*	*Did not occur in project, but a potential disadvantage*	*Occurred, but not regarded as a disadvantage*	*Did not occur and unlikely*	*No Response*
Too much writing for pupils	11 (15·0%)	17 (23·3%)	7 (9·6%)	34 (46·6%)	4 (5·5%)
Too little discussion for pupils	8 (11·0%)	5 (6·8%)	0 (0%)	51 (69·9%)	9 (12·3%)
Lowers standards	3 (4·1%)	9 (12·3%)	0 (0%)	54 (74·0%)	7 (9·6%)
Made it difficult to prepare for O-grade this year	12 (16·4%)	9 (12·3%)	1 (1·4%)	42 (57·5%)	9 (12·3%)
Pupils less willing to work	2 (2·8%)	2 (2·8%)	0 (0%)	50 (68·5%)	19 (26·0%)
Unnecessary anxiety caused to pupils	4 (5·5%)	14 (19·1%)	1 (1·4%)	47 (64·4%)	7 (9·6%)
Over-concentration on SIV by teacher	22 (30·1%)	23 (31·5%)	0 (0%)	25 (34·2%)	3 (4·1%)
Folders difficult to keep in order	17 (23·3%)	14 (19·1%)	4 (5·5%)	34 (46·6%)	4 (5·5%)

100% = 73 teachers.

The items listed in Table 8.4 were not differentiated as 'Disadvantages for Pupils' in the teachers' questionnaire, but were part of a longer list which included the five items in Table 8.5, now labelled 'Disadvantages for Teachers'.

TABLE 8.5: DISADVANTAGES OF FOLIO ASSESSMENT FOR TEACHERS: TEACHERS' VIEWS (1977)

Disadvantages	*Occurred*	*Did not occur in project, but a potential disadvantage*	*Occurred, but not regarded as a disadvantage*	*Did not occur and unlikely*	*No response*
Too much marking (of normal classwork)	21 (28·8%)	15 (20·5%)	10 (13·7%)	22 (30·1%)	5 (6·8%)
Too much planning time required	12 (16·4%)	14 (19·1%)	10 (13·7%)	30 (41·1%)	7 (9·6%)
Too much 'Trial Marking'	33 (45·2%)	15 (20·5%)	10 (13·7%)	11 (15·1%)	4 (5·5%)
Too much administrative work	25 (34·2%)	21 (28·8%)	7 (9·6%)	11 (15·1%)	9 (12·3%)
Problem of finding teaching material if no past papers	30 (41·1%)	15 (20·5%)	12 (16·4%)	11 (15·1%)	5 (6·8%)

100% = 73 teachers.

Only in answering this last group of questions is the majority of the teachers found to regard the folio assessment scheme unfavourably. Concern about their ordinary marking load was heightened for many by the extra obligation they felt the project imposed to pay close attention to *all* the folio work, even though it had been agreed that not all work need be assessed at the time it was done. The fact that a moderator might be looking at the folios probably helped create this feeling, as well as pupil and parent pressure to assess everything. The 'Trial Marking' exercises involved more assessment work and a considerable amount of administrative work, too, all of which, along with the final assessment of the folio and the recording of scores, overloaded the teachers. Many of them felt, too, that the research team's suggestion that they should not use past papers as teaching material added considerably to their burden. *The result of so much pressure of work was that, despite the perceived educational and professional advantages of the scheme, a majority of the 73 respondents said they would not welcome a similar scheme to replace the O-grade examination and more opposed than supported the suggestion that it might be an alternative to the external examination.* The figures are shown in Table 8.6.

TABLE 8.6: DESIRABILITY OF FOLIO ASSESSMENT (1977)

	Yes	*No*	*No response*
1. *Folio to replace O-grade*			
	11 (15·1%)	42 (57·5%)	20 (27·5%)
2. *Folio as alternative to O-grade*			
	29 (39·7%)	30 (41·1%)	14 (19·1%)

100% = 73 teachers.

The questionnaire enquired also about the extent to which the project influenced the course taught by the teachers. Fourteen teachers in five schools said they had not used past papers when they would normally have done so, and a further 15 in these five and other schools set less past paper work then usual. Four teachers omitted some other work they would otherwise have done: in one case this was a test on 'the novel' and the other three, from the same school, said that they had felt an obligation to get pupils writing to fill the folder to the detriment of general language work and oral work. On the other hand, 20 teachers felt they had set new types of work specifically because of the project. This fell into three categories: (1) a wider range of writing tasks and stimuli, (2) a wider range of 'traditional' interpretation passages, (3) more 'close reading' of literary texts.

II. Comments

The views of teachers were obtained several times in the course of the experiment, by invitation to Principal Teachers to write to the research team, and also, informally, during Trial Marking exercises and the researchers' and moderators' visits to the schools. The contents of Tables 8.2-8.5 may be regarded as indicating the range of matters of concern which came to notice in these ways, since almost all the listed advantages and disadvantages were mentioned by the teachers. The answers to the last questionnaire item, inviting free comment, perhaps show which issues some teachers felt strongly enough to expand on. Almost all the statements made are given below: those omitted related to aspects of the project not bearing directly on the desirability of course-work assessment, such as the nature, length and timing of the Criterion Test, or simply repeated a comment made by the same teacher elsewhere on the questionnaire. A few remarks made in answer to questions earlier in the questionnaire are included here with the final comments.

Numerically, the statements can be divided under the following headings. (The groupings are, however, somewhat arbitrary and several comments make more than one point.) While the numbers of favourable and unfavourable statements give a rough indication of the issues causing most concern, it should be remembered that the totals include more than one comment by some particular teachers.

NUMBERS OF FAVOURABLE COMMENTS
RELATING TO:
ASSESSMENT AND COURSE PLANNING – 7
PUPIL MOTIVATION – 6
'TEETHING TROUBLES' – 3
Total – 16

NUMBERS OF UNFAVOURABLE COMMENTS
RELATING TO:
OVERWORK – 11
PAST PAPERS – 6
PUPIL MOTIVATION – 4
FOLDERS – 3
MISCELLANEOUS – 6
Total – 30

Favourable Comments

Assessment and Course-planning

Helpful for internal assessment within the department. (Fife 4)

The experiment proved valuable in that a departmental forum for discussing and justifying marks was set up, and as a result I felt I examined my own assessment of material far more closely than I had ever done previously. (Grampian 4.)

In general I would say that the folio system is fairer to pupils by eliminating the 'lottery element' of the exam system. In addition potential Higher candidates are not restrained by the demands of the O-grade exam. (I know this shouldn't happen, but it does.) (Grampian 4)

Wider range of creative writing tasks and more detailed close reading of areas of texts in literature with written answers. (Work done in 76-77 different from normal.) (Grampian 2)

Top-flight pupils can develop more quickly and are not restricted to O-grade work for most of the year. (Grampian 2)

In general, continuous assessment and internal assessment are worthwhile and could be used as part, at least, of the overall assessment of pupils for O-grade. Teachers of English are made more aware of the aims of an O-grade course and the problems of assessment. A single assessment of the completed folder does, however, seem a bit inadequate. (Grampian 2)

The production of original material (for interpretation etc.) is a potential advantage, if done in an organised system, although it is sometimes hard for individual teachers to ensure a suitable standard of questions etc. (Fife 3)

Pupil Motivation

Main advantage was to provide a useful motivation to less good O-grade classes. (Fife 4)

Increases amount of work done by poorer pupils because of greater motivation. (Fife 4)

Class has improved on an all-round basis of about 20/25%. I feel there has been a greater application to class work since the first assessment. Class has learned to apply itself with greater concentration; language and word sense has been sharpened to a greater degree than previously. The main improvement has been mechanical/technical in the fields of language – greater spelling, punctuation, grammar, expression. I feel this project has helped the class tremendously. (Strathclyde 5)

Advantage that there is a requirement placed on pupils to work through the whole year. (Fife 3)

Classes liked this method of working – felt it fairer. Everyone felt they had a better chance and were not rushed. I myself feel it is a fairer system and would like to see it adopted on a national basis.

> I found this scheme an advantage in that it would force pupils to work throughout the year. I find it satisfactory for those pupils who don't try all year and then manage to 'turn it on' for the exam. Also it helps those who are 'bad examinees'. (Strathclyde 2)

'Teething Troubles'

> I would say that the staff are now reasonably satisfied with the operation of the project. Most of the initial problems and uncertainties have been cleared up and the staff are attuned to the process involved. The stresses and strains of day-to-day teaching can magnify quite minor problems. One of the stresses, of course, is caused by class exams which go on regardless of this project and which inevitably, in the setting, sitting and marking, take up a sizeable chunk of time in each term. (Grampian 4, December 1976)

> I found it a little difficult running this scheme as well as the O-grade. I think that if used on its own it could really work. It was difficult to avoid using past papers because it was necessary to prepare for the O-grade itself. (Strathclyde 4)

> Generally, I approve of the scheme though it does place burdens on the less able pupils and their teachers and there was a great deal of administrative work involved. (Strathclyde 2)

Unfavourable Comments

Overwork

> The amount of staff work implied is huge, given that all teachers will have something like 4 other classes, likely to include one H, and possibly an SYS. Twenty-one substantial corrections in a session for one class must affect the others. N.B. Three of my staff had *two* S4 classes in 1976-77. (Fife 4)

> Moderation and final cross-marking takes vast amount of time. (Fife 4)

> 'I would welcome the introduction of a folio scheme of assessment.' . . . not unless paid for extra marking! (Fife 4)

> Life is too short – particularly when dealing with *two* fourth year classes. Other classes' work suffered. (Fife 4)

> The entire project has been very time consuming and tended to become rather tedious at times when preparation-correction time was at a premium. (Grampian 2)

> Tremendous amount of extra correction and extra paper work. (Fife 4)

> Single assessment of folder required *after* hurried marking of 4th year prelim exams, 3rd year and 2nd year exams. (Grampian 2)

> Minimum requirements. It is clearly not going to be possible to adhere to the minimum requirements with the bottom set. These requirements are already nearly exceeded by the top set. (Grampian 4, December 1976)

Willingness to produce second version or third version increases pupils' workload. (Fife 5)

The main criticism of this particular experiment was the lack of time. More time might have been allowed for the discussion of the scheme before it was introduced and sample materials should have been available to the teachers involved months earlier. (Grampian 2)

The 'Rules for Folio Assessment' agreed last summer did not arrive until late autumn when the session's work was well under way – therefore I paid lip service to the rules and, while writing was well represented, reading and interpretation were under-represented.

I felt my contribution was negligible since the scheme was presented with no time to give it adequate preparation or planning. My group contains a high proportion of potential Higher candidates and I did not find it helpful to be instructed to give a student 30/30 because it was top O-grade standard. The scheme seemed to have two aims: (a) to standardise teacher assessment (b) to produce a wider range of teaching and testing material. These aims would have been better served had they been treated separately. (Grampian 2)

(Note on the last two comments: the criticisms made indicate some of the problems of communication of instructions and advice. In fact, the material referred to was sent to the school in April, for an August start to the experiment, and was discussed at two meetings with the department in April and June.)

Past Papers

Impossible not to use past papers as no substitute is offered or suggested. If teachers were responsible for producing and preparing interpretations etc. of their own they would have no time to teach! Anyway what is wrong with using past papers discriminately? (Grampian 4)

Work done in S4 academic classes must be seen as preparation for O- and/or H-grade exams. Where pupils hope or expect to pass these exams, teachers will be failing in their duty if they do not teach for these exams. Defenders of the Scheme will doubtless claim that it provides adequate preparation for SCE exams, and so in theory it may do. But in practice it cannot, at least as originally envisaged. And using the Scheme only in part or in watered-down form can prove nothing, while it will place an extra burden on teachers in marking and administration. (Grampian 2)

At the moment it is essential for the pupil to succeed in the formal O-grade exam *as well as* having his standards improved. (The former does not necessarily follow from the latter.) Formal examination practice, such as is provided by past papers and Pillans & Wilson, is necessary for this. (Strathclyde 5)

'I fail to see why past papers cannot be used for class practice in addition to other interpretation work. The amount of preparation and correction time sometimes makes it extremely difficult to prepare "questions" for the pupils – thisis a purely practical point!' (Fife 4)

Work omitted because involved in this scheme? – Past paper work, which I feel is very beneficial (since they still have to sit the O-grade exam). (Fife 3)

Fewer interpretations done – difficulty of making up original tests in sufficient quantity. (Fife 3)

Pupil Motivation

Pupil motivation is impossible to gauge. There was in fact the average crop of absentees, disciplinary suspensions, failures to hand in written work which suggests that one awkward element of the population was no more impressed by this than by my usual demands for a fair proportion of written work as a condition of entry. (Fife 5)

The pupils involved did not take the project seriously. The impact of exams was completely lost and much of the written work demanded was not done. It was quite impossible to get any work done at home. (Fife 3)

The scheme in itself could work, but only given pupils who respond. The group I had were, perhaps, the poorest working group I have dealt with. It was almost impossible to make them undertake the assessments set and the standard of work suffered. An exam (with the less motivated pupil) has an impact and means much more than any other viable alternative I have yet seen. (Fife 3)

Lack of Prelim/Prelims lessened motivation of pupils. Folder-work soon lost its importance for poorer/less willing pupils. Lack of timetabled Prelim left pupils short of 'exam' practice. (Fife 3)

Folders

Loss (of folders) due to lack of classroom security. (Strathclyde 4)

Folders make revision difficult. (Fife 5)

Need for jotters. The staff felt it was necessary to provide jotters for rough work and for some exercises. The folders received by the moderator do not include these jotters which do in some cases contain a substantial amount of work. Presumably suitable work in these jotters should be used for assessment work although up till now staff have concentrated on material in folders. (There was no real justification for supplying a jotter other than the teacher's deep-felt belief that a pupil needs a jotter!) (Grampian 4, December 1976)

Miscellaneous

Considerable guidance required for marking – I don't know how. The notes for marking compositions are very helpful, but lengthy. One copy can also easily be lost or mislaid. (Strathclyde 5)

I found that there was greater pressure to produce a stockpile as an end in itself. In the normal year this stockpile emerges as a natural end-product of class work. I felt that my objectivity, my own capacity to assess soberly and therapeutically, was slowly eroded by the endlessness of it. I felt also that the advantages in motivating *some* were offset by the tantrums of others who for the first time actually expressed their frustration and sense of inadequacy with continuing failure.

There is a final conflict: are we assessing entirely independent of O-grade prognostications or are we trying to match ourselves against the O-grade as some kind of sacrosanct absolute? Although the idea is clearly to replace the O-grade, this year's experiment has been bedevilled by the necessity of trying to foretell accurately how the candidates will actually perform: not the same thing as selecting the highlights from a long period of work. (Fife 5)

National examination is preferable. (Fife 3)

I have a dislike of any scheme of internal assessment, as opposed to an outside exam, since it involves too much of the personal. We live here, and meet frequently the parents of our pupils. Examiners are faceless (? frozen kippers) and opprobrium attached to them is acceptable. In the small community, it would not be, despite the intervention of moderators – who, for the most part, are also our friends and colleagues. (Grampian 1)

Single assessment allows no scope for feedback to pupil. Gives teacher no continuing idea of progress. (Grampian 1)

The scheme encouraged individual initiative, but might not have this effect in the long-term – merely providing a set of hoops of an initially different form which might well become as standard as those they replace. The removal of certain 'props' might have an adverse effect by reducing general confidence in one's own teaching – even the reversion to the odd 'Pillans & Wilson', with a threat that it might be used in evidence, has a consoling effect on occasion. (Fife 3)

III. Some Recommendations about Practicalities

Regarding practicalities and teachers' opinions only, it would seem that the following would be desirable elements in any future internal assessment scheme.

1. There should be adequate planning time – perhaps one year.
2. During this time there should be planning of course aims and Trial Marking exercises to 'teach' assessment techniques to the teachers
3. Some ways should be found to reduce the time commitment of teachers when the scheme is in progress:

   e.g., (a) Reduction of the minimum number of pieces of work for assessment from 14 to 9 or 10, say 4 compositions, 4 'responses to Literature' and one or two interpretation tests common to the whole school.

   (b) Schools should have Folio Assessment or Internal Examination, not both (because of the *time* involved).

   (c) Serious consideration should be given to the provision of more free time for English teachers for assessment and/or to the reduction of class sizes to reduce the load of written work.

4. The favourable motivation of teachers should be sought by

   (a) Giving internal assessment significant weight if there is also an external examination.

   (b) A shift of emphasis as to the purpose of assessment in English, to encourage greater concern for description of a pupil's strengths and weaknesses, which is better done by teachers than by external examiners.

   (c) (Ideally) the provision of some payment for internal assessment for certification.

5. Instructions should be clear and simple with regard to
    (a) the purpose of the assessment.
    (b) what is being assessed (i.e., *achievement*, rather than ability).
    (c) the method of recording assessments.

6. Each school involved should delegate responsibility for the implementation of internal assessment to a member of the English department who is committed to the method and has the time, energy, knowledge and authority to encourage others, arrange meetings, give advice on standards within the school, impose deadlines, insist on adherence to instructions and engage all in Trial Marking and Moderation procedures.

CHAPTER XI

CONCLUSIONS

The SCRE project was undertaken in pursuit of the following aims, which are also set out on the first page of this report.

1. To report on the present practice in assessment by schools of candidates for Ordinary Grade English.
2. To investigate and determine the optimum ways in which teachers can make assessments (possibly in the form of orders of merit) of pupil performance in selected aspects of English, these assessments being based on a wide range of work over a substantial portion of the school year.
3. To determine the best ways of scaling and/or moderating these assessments so that assessments of pupils in different schools may be comparable.
4. To investigate such other related matters as may appear in the course of the investigation to be relevant to the field of interest covered by the Board's remit.
5. On completion of the above steps to make recommendations to the Board on whether alternative means, utilising teachers' assessments, should be substituted for the whole or part of the present O-grade examination in English, to outline the administrative changes that would be involved. Costing of these changes might also be undertaken in co-operation with officials of the Board.

The findings of the project cannot easily be reported in direct relation to each of these aims, because, as the research progressed, some of the 'related matters' referred to in Aim 4 were seen to be so significant in the evaluation of internal assessment as compared with external that they assumed major importance. Principal among these significant 'related matters' are: the well-known problem of marker unreliability in all assessment of English; the tendency for markers to fail to spread their marks out widely (which may be a reflection of the nature of English and of the actual distribution of 'English skills' in the population); practical difficulties in implementing internal assessment, caused mainly by severe constraints of time; and the reaction to internal assessment of the English teachers involved, which ranged from enthusiastic commitment through indifference to strong hostility. Conclusions are also influenced by another related question, that of the relationship between assessment procedures and the courses followed by pupils.

The researchers were not breaking new ground in discovering that the assessment of English in Scotland is fraught with the same problems as are met by teachers and examiners elsewhere, and most of what they have to report about their experience bears close resemblance to the findings of other researchers into English assessment, such as the authors of the various Joint Matriculation Board reports on their GCE O level Alternative English Language assessment and the Schools Council reports on CSE assessment.

The SCRE project was in some respects a feasibility study in Scotland for a type of assessment already employed elsewhere fairly successfully and the characteristics of which are widely known. The principal significance of the present report is probably that it brings to light in a Scottish context various aspects and problems of the assessment of English, and serves as a reminder that they apply as forcibly here as elsewhere.

In two respects, however, the researchers believe that they can claim that the SCRE project has been unusual. The presentation of the tables and figures in the statistical evaluation of internal and external assessment has been deliberately designed where possible to show the effect of such factors as marker unreliability and the effects of standardising and/or scaling raw scores on particular pupils; mere comparison of correlation coefficients or of pass rates, it was felt, does not always lead to a clear realisation of the absolute numbers of pupils affected by discrepancies between markers, or shifted across band boundaries by a scaling process. The second unusual element in the research is the testing of the frequent assumption that a school's rank order is likely to discriminate among pupils more validly than that produced by an external examination of normal length: the Criterion Test allowed comparison of both with a third measure which was as extensive, valid and reliable a test of English as could be put together and administered in reasonable time. The process of making the Criterion Test, which involved consideration of English tasks and skills, has incidentally provided material likely to be of value in developing diagnostic or profile assessment in English.

What is the best way of relating the researchers' findings to the formally stated aims of the project? A re-statement of the aims (without distorting them too much) in the form of five broad questions seems helpful.

1. What are the advantages and disadvantages of retaining unchanged the existing O-grade English arrangements?
2. What gains and losses would ensue if O-grade English were assessed wholly internally?
3. What are the advantages of combined internal and external assessment?
4. Which methods of moderating standards are most effective?
5. What should be English teachers' main concerns?

1. *For and against the existing examination*

The O-grade English examination has many advantages. Its administrative arrangements are well-tried and effective, and its nationwide currency ensures comparability of standards for pupils from all schools; in this respect there is particular advantage in having a single common interpretation test. There is a general satisfaction with the O-grade examination among both teachers and other users of its results. It is, indeed, very professionally set and is probably as valid as possible each year. Careful consideration has been given over the years to the 'washback' effect on the school curriculum of the form of the examination, so that, in principle, teachers are free to teach a wide variety of courses preparatory to O-grade. (There is evidence that in practice, however, many teachers do not take advantage of this freedom and make extensive use of O-grade past papers as teaching material in S4.) A further credit to the examination is the opportunity it affords to some candidates to show their real worth, which they may have lacked the motivation to reveal in school work; this benefit is, of course, offset by the disadvantage to pupils who work well in school but are 'bad examinees'.

The more unsatisfactory characteristics of the O-grade examination, besides any undesirable influence on teaching, are threefold:

(a) it is bound to be of limited validity because it can sample only a small proportion of possible English work;

(b) like all assessment of English, it suffers, despite marker-standardisation, from serious inter-marker inconsistency, which probably causes some 25% of candidates to obtain marks a full band away from those they would receive if a different marker marked their answers in Paper I (Composition and Literature);

(c) it is not a fine discriminator, especially among a very large body of 'average' pupils and especially in Paper I (Composition and Literature).

There are no obvious means of reducing the effects of these major disadvantages, short of the probably impractical solution of multiple- or, at least, double-marking to improve marker-reliability. Some improvement in marker-reliability would probably occur if dispersions of marks were standardised, as well as means, but the effect of this would not be very great.

2. *Internal Assessment only*

If assessment for certification at the end of S4 were wholly internal, the main advantages would be gains in validity: a school examination, or, even more so, Folio assessment, can be much more directly related to the courses taken by the pupils than an external test, and Folio assessment covers much more of the pupils' work than an external examination. (Though the SCRE project concentrated on the evaluation of Folio assessment, it is likely that school examinations would, in general, provide an equally good alternative to the O-grade examination.) Chapter X of this report shows that teachers would probably see educational and professional advantages in internal assessment, and would be likely to feel more free in practice to teach a variety of courses and would plan them more carefully. In-service training through Trial Marking exercises would be needed, but the extra work involved would probably be regarded as worthwhile by the teachers. Internal assessment for certification would probably make a stronger emphasis on diagnostic assessment more likely, too.

Numerous problems would, however, arise. While many schools, with appropriate in-service training in assessment, would assess their pupils satisfactorily, a number would fail to do so for a variety of reasons. The setting of valid and comparable interpretation tests in different classes or in different schools presents great difficulties, which affect both school examinations and Folio assessment, especially the latter. Marker-inconsistency would be at least as prevalent as in the O-grade examination, and the schools would be slightly more likely to 'bunch' the marks than O-grade markers, so that they would prove no better discriminators. In addition, there would be some danger of schools' under-marking uncooperative pupils and over-marking some thought to be of high quality because they are in top classes. Moderation presents practical and theoretical problems. Above all, the difficulties for teachers of implementing an internal assessment scheme in current working conditions are considerable. The evidence is in Chapters IX and X. Many teachers would be very reluctant to take on the work involved. Internal assessment could not replace the O-grade examination and effectively carry out its functions in the present circumstances.

3. *Advantages of combined internal and external assessments*

While the improvement is not so marked as when multiple-marking of the same script is carried out, it can be shown that the reliability and discrimination of writing and literature assessments improves when different single-marked assessments are added together. A combination of internal and external marks for these areas of work would in most cases be a better assessment than either alone, provided that both are validly assessing some aspects of English. Even when internal assessment was significantly less satisfactory than external in the SCRE project, the average of the two correlated satisfactorily with the Criterion measure. It is desirable that such combined marks have equal weight because any other arrangement would increase the influence of the marker-unreliability affecting the more heavily weighted assessment. Equal weighting would also ensure teacher- and pupil-motivation. If the school assessment covered a number of different types of writing – say three from 'Expressive, Factual, Persuasive, Fictional Story, Dramatic Dialogue, Poetry' – the external test could consist of a single topic to be dealt with in a specified style or with a specified purpose and audience in mind: this procedure would improve the reliability of the external marking but would not have a detrimental effect on teaching, since the pupils would have to engage in several kinds of writing for the internal assessment, The topic and style specified in the external test could vary from year to year,

An arrangement such as is suggested would improve the reliability of the external assessment and provide a more satisfactory rank order upon which the SCEEB examiners could fix band boundaries according to their judgment of standards. Any consistent departure by the schools in the internal assessment from SCEEB standards could be corrected by scaling internal against external marks, while the combination of the two marks would reduce the effects of inter-marker inconsistency. It would, however, be necessary to carry out some moderation of assignments set in the schools to ensure their validity. This might be combined with Trial Marking, or effected by schools sending copies of assignments to SCEEB or to another appropriate body.

The researchers believe that at present it would be desirable to retain some assessment of interpretation in the hands of the external examiners, because of the lack of comparability between interpretation tests in different schools. The SCRE project has not, however, studied the assessment of interpretation as closely as that of writing, and it may be that a thorough investigation of the validity and reliability of the traditional interpretation test would show that it, too, has more faults than is commonly realised. The whole question of the relative weightings of different elements in S4 English assessment might need reconsideration as a result of such an investigation, perhaps so that greater weight was given to the response to literature, or so that the assessment of kinds of reading other than literature and interpretation passages acquired some significance. Schools could certainly devise alternative ways of testing reading: so long as it is necessary to produce a school rank order, however, it will be desirable that there be at least one test common to all classes, with a common standard of marking.

The combined assessment scheme suggested would, in principle, allow a reduction in the size of the external examination, since the writing task could be shorter (at least when certain types of task were set) and it would be possible to leave the assessment of literature wholly to the schools, if one was prepared to accept the disadvantage that it would be necessary to scale internal writing and literature against external writing and interpretation marks. Depending on success in developing them and their acceptance by English teachers, multiple-choice tests of interpretation might be employed, which would reduce marking costs. The SCRE data has indicated that appropriately weighted combinations of writing and either traditional or multiple-choice interpretation tests cover fairly satisfactorily the range of skills also tested by the Folio Assessment.

The practical problems of internal assessment would still apply to a combined scheme, but would be more manageable without the need to set interpretation tests on the O-grade pattern or to prepare pupils for an external literature test. The researchers believe that the professional advantages to teachers in the acquisition of knowledge about, and practice in, assessment, along with the improvement in validity and reliability such a system would probably bring, outweigh the difficulties it would meet. Nevertheless, it would be advisable to reduce the burden of assessment carried by teachers by requiring internal assessment of about ten pieces of work, rather than the fourteen or fifteen demanded by the SCRE project.

4. *Moderation*

The researchers' views on moderation of internal assessment are implicit in the preceding discussion. Some moderation of courses and assignments would be necessary and would probably be welcomed by the teachers as helping them to plan their work better and providing them with new ideas and suggestions.

As to moderation of standards, the judgment of the SCEEB examiners is the touchstone, and so long as it is important to effect the kind of discrimination among pupils which the existing examination achieves, the most effective way of applying the touchstone to a large number of internal assessments is probably by the use of a reference test such as the external writing and interpretation test mentioned in 3, above. Scaling of internal marks might, however, be more helpfully carried out against raw external marks, to avoid the exaggeration of discrepancies on ranking which occurs if scaling is against external marks standardised in the SCEEB pattern.

There is a need for English teachers and examination markers to be made aware that reliable fine discrimination is not possible between one pupil's 'performance' in English and another's, especially within a large mass of 'average' pupils. It would be preferable for markers and teachers to think of discrimination in terms of 5 bands rather than in terms of 14 ranges or 50 or 100 marks.

5. *Teachers' priorities?*

The researchers' experience in grappling with the problems of discriminatory assessment did also lead them to consideration of the appropriateness in English of placing so much emphasis on this purpose of assessment as opposed to others.

Future development in the assessment of English in S4 depends ultimately on value judgments about priorities among purposes of assessment and about the degree of control over curriculum teachers should have. The scheme put forward in the booklet on assessment published by the Australian Council for Educational Research (ACER, 1974), quoted in 'Internal Assessment or External Exam?' (Spencer, 1975), puts the matter neatly.

The article stated:

> 'Methods of assessment can be placed on a continuum representing (a) increasing levels of public trust and confidence in teachers, (b) decreasing concern for discrimination and competitiveness and increasing desire to have a description of the pupil's achievements rather than a statement of the quality of his performance relative to others in his group. The nearer an examining body wishes to be to the left of this continuum the more important is standardisation of any internal assessments; the nearer the Board is willing to be to the right of it the more need there is for teachers to be trained in curriculum development and evaluation.'

External exam at end of year or course	Various types of moderated school assessment	Assessment without competition

→

Increasing levels of public trust in teachers

←→

Norm-based exam results on a certificate. Concern for discrimination.	Report of pupil's achievements. Concern for diagnosis and feedback to pupil and teacher.

Realisation of the inadequacy of discriminatory assessment in English could have fruitful results for other sorts of assessment. The general reappraisal of assessment in S4 occasioned by the proposals of the Dunning Committee provides a context in which a shift in emphasis in the assessment of English could occur. The in-service assessment training recommended by that Committee should concentrate first on assessment design. Course-planning takes a primary place in the process, and the assessment is designed to match the purposes with which the teacher taught and the pupils learned. English departments need to develop consciousness of all the possible elements of the English curriculum suitable for their particular pupils and then, in planning a term's or a year's course, to make clear choices among them so that they can specify (yet without inflexibility) the purposes for which they and their pupils will work in the immediate future. When the assessment is closely related to particular purposes in this way it is possible to obtain a more specific description of achievements in English than if the principal concern is to discriminate among the pupils. Assessment of this sort can result in the diagnosis of strengths and weaknesses or in the noting of attempts by the pupil at usages which show that his linguistic awareness is developing – outcomes much more closely related to the business of teaching and learning than are grades achieved in a discriminatory examination.

Efforts to seek comparability of standards should not be abandoned, but the state of English testing and teaching might be the better for a more general awareness that it is not in practice attainable, even in a good external examination which takes careful measures to ensure as much fairness to all candidates as possible. If this knowledge permeated the consciousness of English teachers at the same time as they were given a significant weight of assessment to carry out themselves, the powerful influence of the external examination on what is taught in S 4 would be weakened, more varied courses could then be developed in more schools, and the quality of those courses could be more immediately and more effectively evaluated by the teachers who designed and taught them.

Appendix 1

Questionnaires Used in Survey of English Teachers' Views, 1975

(see Chapter I)

Appendix I(a): Free Response Questionnaire

This is not included in this publication but is available as a separate item, on request, from the SCE Examination Board.

N.B. The 'Objectives of the Ordinary grade assessment of English' which the teachers commented on can be seen in a very slightly different version in Appendix 3, page 111.

Appendix I(b): Fixed Response Questionnaire (with results)

Name of school ....

Please tick the relevant boxes:

Position in English Department:

1. Principal Teacher	(____)	89	17%
2. Assistant Principal Teacher	(____)	79	15·1%
3. Assistant	(____)	355	67·9%
		523	100%

No. of years teaching English:

4. 0 – 2	(____)	111	21·2%
5. 3 – 5	(____)	126	24·1%
6. 6 – 8	(____)	65	12·4%
7. 9 – 15	(____)	102	19·5%
8. 16 or more	(____)	119	22·8%
		523	100%

NOTES

1. *Definition of Objectives.* Speaking, reading, writing, and the assessment of success in these activities, are complex functions involving sets of skills, personal characteristics and experience. Most pieces of writing, for instance, will show that pupils have achieved objectives in several of the categories set out in this questionnaire. The classification used will, it is hoped, clarify what 'performance' in English involves, but it would not be desirable, or possible, to assess the achievement of each objective separately. Further, the list is not comprehensive. Some factors affecting English 'performance', such as the stage of intellectual and emotional development a pupil has reached, cannot be accurately measured with existing testing tools. There are also, even at O-grade standard, qualities recognised as intrinsic to good

communication – e.g., breadth and aptness of vocabulary and variety and appropriateness of sentence structure – for which specific criteria cannot be formulated in advance. The purpose of the questionnaire is therefore to obtain an overall impression of the collective state of mind of English teachers relative to objectives which can be specified with some precision, but the research team remains aware that not all the qualities of 'good English' are definable in advance.

2. *Classification*. Other kinds of classification of types of writing and objectives are possible. (The one used here is derived directly from James Britton's 'What's the Use?' in *Language in Education*, published for the Open University by Routledge and Kegan Paul.) The research team will welcome comments from any teachers who can suggest improvements or a superior type of classification. It is, however, essential in order to maintain comparability among all the respondents that the questionnaire be filled in as it stands.

3. *Standard*. It will help throughout the questionnaire to keep in mind the O-grade standard, which is defined as being 'such that a pupil who is at the lower end of the top 30 per cent of an age group should, with satisfactory teaching and adequate effort on his part, have a reasonable prospect of securing passes (i.e., Grade C) . . . in at least *three* subjects in the fourth year'. ('Report of the Working Party on the Curriculum of the Senior Secondary School', HMSO, Edinburgh, 1959.) There are no figures available to indicate the proportion of an age group expected to obtain Grade C in English alone, but a reasonable estimate is that, in the conditions mentioned above, *a pupil at the lower end of the top 50 per cent of an age group is likely to perform in the fourth year at about the Grade C/D borderline in English alone*.

4. '*Jargon*'. Technical language from various fields of study has been employed in the questionnaire for convenience and in pursuit of precision: it is not implied that it should be used in the classroom.

Part One: Objectives

WRITING

Three main headings are proposed:

(*a*) 'Transactional' writing: to direct, question, get things done, participate in society . . .

(*b*) 'Expressive' writing: to exchange or reveal feelings and opinions, convey attitudes, reveal personality . . .

(*c*) 'Poetic' writing: to construct linguistic artefacts – organised language – as a means of trying to explore and master the complexities of reality.

Please tick the appropriate boxes to indicate which objectives under these general headings are, in your opinion, the *minimum* a pupil should be capable of achieving to obtain (i) Grade C, (ii) Grade A in an assessment of O-grade English. (Some may be thought more difficult than the minimum requirements for Grade A.)

'TRANSACTIONAL' WRITING

Pupils should be able to:

		Grade C	*Grade A*	*No. not choosing item*
9.	write instructions for a simple task (recipe, care of sports equipment . . .)	498 95·2%	10 1·9%	15 2·9%
10.	formulate rules for a club, social evening, or some situation in which group behaviour must be controlled	264 50·5%	193 36·9%	66 12·6%
11.	report events factually	488 93·3%	24 4·6%	11 2·1%
12.	write notes on events as they happen	312 59·7%	141 27·0%	70 13·4%
13.	write a report presenting information logically	255 48·8%	235 44·9%	33 6·3%
14.	order and paragraph ideas	439 83·9%	69 13·2%	15 2·9%
15.	write an argument based on facts for or against a point of view	261 49·9%	245 46·8%	17 3·3%
16.	present evidence and draw conclusions from it	131 25·0%	341 65·2%	51 9·8%
17.	give arguments for and against a point of view and draw rational conclusions	95 18·2%	348 66·5%	80 15·3%
18.	illustrate a general idea with a particular example	263 50·3%	213 40·7%	47 9·0%
19.	develop an idea, illustrating, quoting, adducing evidence in support	62 11·9%	354 67·7%	107 20·5%
20.	write a logical, theoretical argument, given premisses	22 4·2%	302 57·7%	199 38·0%
21.	write advertising copy (e.g. for school events . . .)	350 66·9%	124 23·7%	49 9·4%
22.	write a persuasive argument using emotive language to win support	56 10·7%	360 68·8%	107 20·5%
23.	write a speech for or against a motion	262 50·1 %	225 43·0%	36 6·9%
24.	summarise factual information	445 85·1%	67 12·8%	11 2·1%
25.	summarise an argument, maintaining the logical thread	171 32·7%	291 55·6%	61 11·7%
26.	write personal letters of various kinds	497 95·0%	16 3·1%	10 1·9%
27.	write formal letters to achieve practical ends	462 88·3%	49 9·4%	12 2·3%

'EXPRESSIVE' WRITING

Pupils should be able to:

		Grade C	*Grade A*	*No. not choosing item*
28.	describe their own feelings about and, as stimulated by, events they have experienced	479 91·6%	35 6·7%	9 1·7%
29.	define their attitudes to aspects of society which impinge upon them (e.g. authority, other groups, other races)	279 53·3%	205 39·2%	39 7·5%
30.	state opinions on controversial topics	309 59·1%	177 33·8%	37 7·1%
31.	show awareness of the causes of their feelings and attitudes	104 19·9%	296 56·6%	123 23·5%
32.	convey their interest in various topics by writing knowledgeably about them	403 77·1%	105 20·1%	15 2·9%
33.	write fantasy stories (invented situations)	336 64·2%	129 24·7%	58 11·1%
34.	write imaginatively (i.e. with originality) or critically expressing a personal view given an artistic or natural stimulus (film, poem, music, picture, physical experience . . .)	177 33·8%	283 54·1%	63 12·0%

‘POETIC’ WRITING

Pupils should be able to:

		Grade C	*Grade A*	*No. not choosing item*
35.	write stories with some exposition of the causes and effects of events	264 50·5%	196 37·5%	63 12·0%
36.	organise events in a story to create a climax	278 53·2%	215 41·1%	30 5·7%
37.	create suspense in a story	172 32·9%	277 53·0%	74 14·1%
38.	write stories in which the feelings of more than one character are revealed	152 29·1%	283 54·1%	88 16·8%
39.	write stories with genuine interplay between events and characters	63 12·0%	299 57·2%	161 30·8%
40.	reveal character in a story through dialogue	118 22·6%	295 56·4%	110 21·0%
41.	dramatise (not just describe) events in a story	100 19·1%	273 52·2%	150 28·7%
42.	write a story with a moral or message	228 43·6%	189 36·1%	106 20·3%
43.	write adopting the point of view of someone with feelings and attitudes different from their own	110 21·0%	277 53·0%	136 26·0%
44.	write with different degrees of formality to suit circumstances	174 33·3%	259 49·5%	90 17·2%
45.	write in simple poetic forms (e.g. ballad, haiku, free verse . . .)	149 28·5%	182 34·8%	192 36·7%
46.	write in stricter poetic forms (e.g. blank verse, sonnet)	2 0·4%	156 29·8%	365 69·8%
47.	exploit (in prose or verse) the emotive overtones of words	23 4·4%	224 42·8%	276 52·8%
48.	use language figuratively	149 28·5%	287 54·9%	87 16·6%
49.	exploit the sound and rhythms of language to make a stronger impression on the reader	43 8·2%	224 42·8%	256 48·9%

Knowledge of appropriate vocabulary and of a variety of sentence structures is inherent in the objectives listed above and the skill with which this knowledge is employed would be assessed in judging the standard of any piece of writing by comparing it with other work. ‘Technical correctness’ would also be taken into account. While it is not practical to specify in detail minimum criteria in grammar and spelling, some indication of teachers’ views about formal correctness may be obtained from the following section of the questionnaire.

Please tick the appropriate boxes to indicate your view of the minimum requirements at each grade. Pupils in their normal writing should be able to:

		Grade C	*Grade A*	*No. not choosing item*
50.	use full stops correctly in nearly every instance	499 95·4%	19 3·6%	5 1·0%
51.	use the question mark correctly when required	501 95·8%	14 2·7%	8 1·5%
52.	use quotation marks for Direct Speech	496 94·8%	21 4·0%	6 1·1%
53.	use the comma correctly to mark off subordinate clauses and parenthetical or appositional phrases	284 54·3%	215 41·1%	24 4·6%
54.	use the comma correctly in combination with quotation marks	391 74·8%	101 19·3%	31 5·9%
55.	use the exclamation mark correctly	394 75·3%	104 19·9%	25 4·8%
56.	use the colon correctly	118 22·6%	288 55·1%	117 22·4%
57.	use the semi-colon correctly	118 22·6%	298 57·0%	107 20·5%
58.	use the apostrophe correctly	454 86·8%	58 11·1%	11 2·1%
59.	use quotation marks correctly for slang, foreign phrases . . .	268 51·2%	197 37·7%	58 11·1%
60.	use parentheses correctly	148 28·3%	277 53·0%	98 18·7%
61.	use the dash correctly	125 23·9%	285 54·5%	113 21·6%
62.	Consistently accurate spelling should be a prerequisite for the award of	174 33·3%	263 50·3%	86 16·4%
63.	Pupils should regularly show evidence of skill with tenses, subordinate clauses, noun phrases, appropriate linking words and other aspects of grammar contributing to the complexity and variety of sentence structures to obtain . . .	107 20·5%	343 65·6%	73 14·0%

READING

Please indicate as before, your view of the minimum requirements at each grade by ticking the appropriate boxes.

As regards their reading, in or out of school, pupils should be able to:

		Grade C	*Grade A*	*No. not choosing item*
64.	reproduce the gist of a narrative	507 96·9%	8 1·5%	8 1·5%
65.	state their own emotional reaction to the content of what they have read	428 81·8%	79 15·1%	16 3·1%
66.	recognise and state the writer's attitude to his subject, and some of his ideas	286 54·7%	222 42·4%	15 2·9%
67.	state the tone/feeling of what they have read	168 32·1%	306 58·5%	49 9·4%
68.	state their emotional and/or moral reaction to the writer's treatment of his subject	200 38·2%	250 47·8%	73 14·0%
69.	recognise and comment on the accuracy with which literature reflects human characteristics, relationships and problems	87 16·6%	278 53·2%	158 30·2%
70.	recognise and refer to the writer's techniques (e.g. devices for creating suspense, revealing character, or structurally relating parts to the whole, his use of figurative language, or other means he uses to create an emotional and/or moral impression on the reader)	84 16·1%	345 66·0%	94 18·0%
71.	give reasons for enjoying what they have read in terms of the writer's skill	196 37·5%	268 51·2%	59 11·3%
72.	give reasons for the judgement that one piece of writing is superior to another	122 23·3%	278 53·2%	123 23·5%
73.	locate specific information in a book using Contents, Index, Chapter headings etc.	424 81·1%	58 11·1%	41 7·8%
74.	'skim' read to get a general impression of a passage	182 34·8%	242 46·3%	99 18·9%

COMPREHENSION

When carrying out 'close reading' of a piece of writing, pupils should be able:

		Grade C	*Grade A*	*No. not choosing item*
75.	to recognise or recall information explicitly stated	492 94·1%	19 3·6%	12 2·3%
76.	paraphrase or translate such information	426 81·5%	81 15·5%	16 3·1%
77.	make inferences based on information explicitly stated and on their own experience	266 50·9%	221 42·3%	36 6·9%
78.	recognise or state the main points presented, as opposed to digressive, illustrative or secondary ones	389 74·4%	120 22·9%	14 2·7%
79.	reorganise and paraphrase (i.e. summarise the main ideas presented	404 77·2%	105 20·1%	14 2·7%
80.	make a judgment based on their own knowledge, experience and values about the validity of ideas in a text	128 24·5%	269 51·4%	126 24·1%
81.	distinguish factual writing from opinion or persuasion	336 64·2%	164 31·4%	23 4·4%
82.	state the effect on meaning or tone in context of the use of different registers of language	90 17·2%	327 62·5%	106 20·3%
83.	state the effects in context of particular sentence structures	84 16·1%	319 61·0%	120 22·9%
84.	state the effects in context of particular usages of grammar and punctuation	135 25·8%	289 55·3%	99 18·9%
85.	explain the implications, overtones, connotations in context of language used figuratively	83 15·9%	323 61·8%	117 22·4%
86.	state the general idea represented by a particular incident, detail or image	268 51·2%	201 38·4%	54 10·3%

Pupils' knowledge of language ought to include the following concepts: even though they may not be able to use the terminology correctly, they should be able to explain the function of each in context.

	Grade C	*Grade A*	*No. not choosing item*
87. Noun	502 96·0%	12 2·3%	9 1·7%
88. Verb	503 96·2%	11 2·1%	9 1·7%
89. Adjective	502 96·0%	12 2·3%	9 1·7%
90. Adverb	476 91·0%	29 5·5%	18 3·9%
91. Pronoun	466 89·1%	36 6·9%	21 4·0%
92. Sentence	495 94·6%	18 3·4%	10 1·9%
93. Clause	323 61·8%	131 2·5%	69 13·2%
94. Adverbial clause	123 23·5%	225 43·0%	175 33·5%
95. Adjectival clause	124 23·7%	226 43·2%	173 33·1%
96. Noun clause	106 20·3%	228 43·6%	189 36·1%
97. Phrase	418 79·9%	69 13·2%	36 6·9%
98. Subject	415 79·5%	52 9·9%	55 10·5%
99. Object/Complement	334 63·9%	96 18·9%	93 17·8%
100. Singular/Plural	477 91·2%	17 3·3%	29 5·5%
101. Tenses	470 89·9%	33 6·3%	20 3·8%
102. Past and Present Participles	236 45·1%	181 34·6%	106 20·3%
103. Finite/Infinite moods	108 20·7%	212 40·5%	203 38·8%
104. Subjunctive mood	48 9·2%	213 40·7%	262 50·1%
105. Active and Passive voices	186 35·6%	199 38·0%	138 26·4%

109

ORAL ENGLISH

Detailed objectives have not been worked out for this aspect of English work, but it will help to know teachers' views about the need for development of it.

	Agree	*Disagree*	*No response*
106. There is a need for Oral English courses to be planned	377 72·1%	109 20·8%	37 7·1%
107. If it is to be assessed, it is better to assess oral performance completely informally, without a test	316 60·4%	162 31·0%	45 8·6%
108. There is no need to try to assess oral English	123 23·5%	324 62·0%	76 14·5%

Part Two: Methods of Assessment

Please tick one box for each item to indicate your opinion of the usefulness of the existing O-grade English examination for the following purposes.

	Effective	*Of some use*	*Not very effective*	*Probably useless*	*No response*
109. to give to employers and those responsible for Further Education useful information for their selection processes	90 17·2%	345 66·0%	65 12·4%	22 4·2%	1 0·2%
110. to provide information for the teacher about the quality of the course he has taught	49 9·4%	259 49·5%	133 25·4%	74 14·1%	8 1·5%
111. to provide pupils with information about their progress and achievements	80 15·3%	318 60·8%	94 18·0%	22 4·2%	9 1·7%
112. to motivate pupils to work harder	119 22·8%	268 51·2%	92 17·6%	33 6·3%	11 2·1%

Please tick the relevant box to provide information about the following:

	Yes	*No*	*No response*
113. This school is using a system of continuous assessment of English, or the assessment of a folio of work, in some or all of forms S1-S3	305 58·3%	193 36·9%	25 4·8%
114. This school is using a system of continuous assessment of English, or the assessment of a folio of work, in some or all of forms S4-S6	113 21·6%	359 68·6%	51 9·8%
115. I would welcome the introduction of some form of suitably moderated school-based assessment of English at O-grade	266 50·9%	219 41·9%	38 7·2%

Part Three: Comments

Please feel free to comment on any aspect of this questionnaire: the research team will welcome any criticisms, suggestions or statements of opinion.

Appendix I(c): Questionnaire — Methods of Assessment Used in Drawing Up Order of Merit Lists for Candidates Sitting the 1975 English O-grade Examination

This is not included in this publication but is available as a separate item, on request, from the SCE Examination Board.

Appendix 2

Consultants who Contributed to the Making of the Criterion Test

Mr A. D. Buthlay, Senior Lecturer in English, Aberdeen College of Education

Mr R. S. Fyall, The High School, Dundee

Mrs E. Grainger, Glasgow

Mr I. G. Mathieson, Principal Teacher of English, Whitfield High School, Dundee

Mr D. Menzies, Adviser in English, Strathclyde Region, Lanark Division

Sister B. Molloy, formerly Assistant Principal Teacher of English, Lawside Academy, Dundee

Mr M. J. Morris, Senior Lecturer in English, Bell College of Further Education, Hamilton

Mr J. P. O'Neill, Senior Lecturer in English, Jordanhill College of Education

Mrs B. Ramsay, Edinburgh

Mr T. A. Sillars, Depute Head Teacher, Auchenharvie Academy, Stevenston, Ayrshire (now Headteacher, Thomas Muir High School, Bishopbriggs)

A particularly significant contribution to Paper II was made by **Mr J. Inglis,** formerly Principal Lecturer in English, Jordanhill College of Education.

Thanks are expressed also to the many headteachers and teachers who allowed the researchers to pre-test parts of the Criterion Test in their schools.

Appendix 3

Stages in the Development of a Criterion Test

I. Preparatory Work: Objectives, List of Skills, Suitable Materials and Various Constraints

(a) Definition of Terminology
The remit for the researchers referred to 'objectives, as stated by the SCEEB, of the Ordinary grade assessment of English as at present constituted', and also to 'other objectives, if any, of the study of English in secondary schools which are not currently assessed'. A third relevant reference, in the description of the proposed 'comprehensive testing procedure', was to 'all aspects of the performance of pupils studying English in Secondary 4 preparatory for presentation for O-grades'. Despite some ambiguity, it seems fairly clear that teaching or learning objectives were meant, and not assessment objectives, such as, for instance, discrimination among various groups of pupils. The 'objectives' were, however, linked to the O-grade examination, so that some confusion with the criteria of success in that test was possible.

The use of the term 'objectives' in the remit created conceptual difficulties which were actually only encountered and removed as work progressed, but which, for the sake of clarity, are discussed here.

Were teachers' *real* or *stated* teaching objectives under study? How were objectives to be identified? Investigation of teaching and learning objectives as evidenced by classroom practice and pupils' behaviour was not a task which could have been effectively carried out within the time limits of the project without several more researchers. Should there be an attempt to apply to English some version of Bloom's *Taxonomy of Educational Objectives*? Apart from the fact that many English teachers would decry such an exercise as likely to oversimplify the subject and, by its over-explicitness, inhibit teachers' curricular adventurousness, there were theoretical objections to an 'objectives' approach to English, arising from Lawrence Stenhouse's consideration of the nature of work in social and aesthetic subjects.[1] His argument implies that in teaching English, one does not first define aims and objectives and then choose content and methodology to attain them, rather, the *content* of the work is central: English consists of the exploration in language of 'content', whether it be texts, experience, reactions to stimuli, emotions . . . and, far from progressing towards predictable ends, one hopes that some of the results will be surprising to pupil and teacher. The English teacher's business, in Stenhouse's view, is to choose content which has value and is appropriate for his pupils and to present it for exploration, not with defined objectives in mind, but with 'purposes' flexible enough to adapt to unexpected developments.

The author, who conducted the early part of the work alone, interpreted the function of the project as more pragmatic than a theoretical consideration of the nature of English as a subject. He was in sympathy with Stenhouse's thesis and took the view that any exhaustive taxonomy of objectives, after Bloom, would, in any case, be based on observation and description of skills actually exhibited by people using English. The instruction of the remit to 'note the objectives . . . of the Ordinary Grade' and to 'list . . . other objectives' was therefore taken to mean something which did seem within the possibilities defined by the size and time limits of the project, namely, to describe *the achievements for which credit is given* to pupils by the O-grade examiners and by teachers in assessing any additional work done in school. The emphasis fell on 'what you give marks or credit for', and broad teaching aims were inferred from the criteria of success. The difficulty of working in any other way is, incidentally, one reason why assessment systems so powerfully affect English syllabuses.

---

[1] Stenhouse (1970-71)

The term 'objectives' was retained for the two questionnaires on English achievements; it was not explicitly defined, but was obviously not being used in a strict behaviourist sense. Later in the work of the project the expression 'List of Skills' seemed more accurate and was more acceptable to most English teachers, since it does not carry the scientific or deterministic overtones sometimes associated with Bloom's use of the word 'objectives'.

(*b*) *Strategy and Practical Limitations*

The first plan was to produce a 'comprehensive testing procedure', or Criterion Test, by creating test assignments and questions derived directly from, and covering all of, a full list of the English skills which attract credit in the O-grade or in school assessments. The list was to be compiled after consideration of the stated aims of the SCEEB examination, analysis of both questions actually set and marking criteria applied, and consultation with examiners and teachers.

Essentially, this was, indeed, the way by which the final version of the Criterion Test was arrived at, but some snags were met en route. There were three sorts of problem. First, it became clear that a Criterion Test covering the whole of a comprehensive list of skills would last about nineteen hours. Secondly, certain aspects of English work, included in the description of all possible S4 work, were not, in fact, being undertaken by any of the schools which by this time had agreed to participate in the experimental internal assessment scheme. There was no point in devising a test of pupils' ability to recognise slanted press reports, for instance, if none of the pupils concerned had been taught the skill. Further, because the schools were still taking the O-grade examination, teachers were concerned that pupils should be familiar with its format. The question of the assessment of oral English in the Criterion Test was an associated problem. It also would have entailed unfamiliar assessment procedures. In the event, after the researchers had made some attempts to develop and pre-test tests of oral English, the Steering Committee of the project decided that there should be no oral assessment in the Criterion Test because of the special difficulties of standardising it and because it was 'straying too far' from the purposes of the present O-grade examination. The third difficulty was in finding test material – writing stimuli, passages for interpretation, literature questions – which covered the list of skills, were not too unfamiliar to the pupils, and were such that they allowed pupils across a wide range of ability to show what they could do in English. Material chosen because it seemed to facilitate the testing of listed skills often met objections from the consultants on the grounds that it failed to meet one of the other requirements for a good examination question, or that, in the case of an interpretation passage, it lacked intrinsic value as a piece of writing, or would be no more than a meaningless exercise to the candidate, in the case of a writing assignment. Pre-testing of intended Criterion Test elements also led to the rejection of several because they proved too easy, too hard, or too incomprehensible to the pupils.

The Criterion Test as eventually set was, therefore, the outcome of compromises. It was decided that two days was as much as could reasonably be asked of schools for the administration of this test, so it was restricted to about 10 hours' duration. To allay the anxiety of teachers about examination practice for O-grade, the first two papers of the Criterion Test were to take the same form as the O-grade examination. The assignments actually employed in the test were chosen because, of all those pre-tested and discussed with the consultants, they best met the criteria of interest and appropriateness for an examination over a wide ability range and at the same time seemed to test satisfactorily those skills selected for inclusion in the no longer *fully* comprehensive Criterion Test.

(*c*) *Lists of Skills*

As can be inferred from the account of constraints on the size and form of the Criterion Test in (*b*) above, not all of the total list of skills produced by the researchers, nor all the insights gained in the process of compiling the list, were actually employed in the making of the test. The total list and description of the various stages through which it passed may, however, be valuable for other purposes. The analysis of English carried out by the research team was a pragmatic one and did not pretend to be exhaustive, to draw fine distinctions between degrees of sophistication with which skills might be employed, or to lead to a statement of the nature and content of English as a subject based on any logical, sociological or psychological principles of organisation. It might, though, be regarded as having provided for some aspects of English work 'check lists', which may be helpful in developing assessment for diagnosis of strengths and weaknesses, or for producing 'profile' descriptions of pupils' achievements in English. Perhaps it is only through the refinement of such pragmatic descriptions of English skills that one may succeed in producing a comprehensive definition of English as a discipline, or provide more precisely worded criteria for assessment.

The process of defining O-grade English skills began with a consideration of the 'syllabus' of O-grade work laid down in the SCEEB's *Conditions and Arrangements, 1975*. This 'syllabus' statement had not varied for some years prior to 1975. It was noted that the statement was of a general nature and open to different interpretations. To obtain a list of more detailed criteria and aims, a study was made of the questions set in the O-grade English papers from 1969-74, and of the Instructions to Markers for 1973 and 1974. The latter varied from year to year only insofar as was necessary to meet the demands of particular assignments set: as far as Paper I, Composition and Reading, was concerned, the basic criteria of assessment had remained the same since the introduction of the O-grade examination in 1965. When this study was completed, the Principal Examiner and one of his senior colleagues were consulted. They confirmed that the set of criteria and aims in 'Objectives of the Ordinary Grade assessment of English as at present constituted' was a full statement of factors taken into account by the O-grade examiners. The list was then sent to a sample of teachers for their comments, and they, too, were generally satisfied that it fairly and comprehensively described O-grade achievements. (The teachers' comments are reported more fully in Chapter I.)

Under each heading, Composition, Reading, Interpretation, the opening statement is a summary of the purposes of the Paper as set out in *Conditions and Arrangements, 1975*; 'Criteria' were derived from the analysis of questions and marking instructions and from the oral comments of the examiners; 'Teaching Aims' were extrapolated from the criteria by the author.

First List of Skills

Objectives of the Ordinary grade assessment of English as at present constituted

COMPOSITION

To test ability to write interesting and coherent continuous prose.

Criteria

The criteria for assessing this ability are stated or implied in the SCEEB's 'Guidance for Teachers on the Marking of English Essays, Ordinary and Higher Grade'. While the emphasis is on the total impression made by the sets of inextricably linked skills which produce the writing, the following points appear to be considered by the examiners:

1. Relevance, quality, depth and quantity of ideas and observations.
2. Degree to which the 'personality' and 'forcefulness' of the writer are revealed in the composition.
3. Degree of sincere interest in the subject.
4. Degree of originality, 'freshness' of approach to it.
5. Degree to which flow of ideas is unimpeded by faulty punctuation, grammar and spelling.
6. Aptness and breadth of vocabulary.
7. Ordering and paragraphing of ideas, including appropriate use of illustration.
8. Variety of sentence structure, and suitability of choice of structure to the expression of ideas and feelings.
9. Appropriateness of the register of language used to its context.

Certain question choices seem to require also:

10. Ability to write in a specialised style (letter, newspaper article, report . . .)
11. The presence of a greater interpretative and imaginative power than the other topics need.
    (Questions using poetry quotations as stimuli.)

Teaching Aims

The teaching aims which these criteria imply are:

1. That the candidate should have knowledge of various aspects of life, including his own and others' attitudes and feelings and their causes.
2. That his knowledge should have grown out of genuine interest, or personal experience, and out of maturity.
3. That the candidate should possess the confidence to express this knowledge.
4. That he should have enough experience of the uses of language to have acquired the means of ordering it and communicating it fully in writing.

READING

To test how 'well informed' candidates are about their reading, both school work and private reading. In poetry questions, to test how far the candidate appreciates the poet's presentation of the subject, as well as the content of the poem.

Criteria

Consideration of the questions set in years 1968-74 reveals that pupils are expected:

1. to be able to reproduce the gist of a narrative;
2. to be able to recognise and state a writer's attitude to his subject;
3. to be able to state a personal emotional or moral reaction to the writer's treatment of his subject;
4. (in some questions) to be able to compare and make a value judgment between two different treatments of the same subject;
5. to be able to recognise and comment on the accuracy with which literature reflects human characteristics, relations and problems, e.g. the qualities and faults of characters;
6. to be able to recognise and comment on the elements of literature which contribute to its enjoyment, e.g. originality, humour, suspense, setting, theme, plot, images, dialogue, etc. (This is expected in some other questions, besides the poetry ones.).

Teaching Aims

The implied teaching aims are:

1. that reading should be wide and based on personal interest;
2. that pupils should become experienced in recognising and stating their own reactions to attitudes and ideas presented by writers;
3. that they should be made aware of some of the techniques used by writers to create impressions.

INTERPRETATION AND LANGUAGE

To test the candidates' understanding of the content of a passage of literary merit on a simple level and also his understanding of deeper or less immediate aspects of the writer's meaning. Also, to test knowledge of language, and the comprehension of the whole or a substantial part of another passage of straightforward prose.

Criteria

The 1968-74 papers have required candidates –

1. to exhibit a breadth of vocabulary, by recognising meanings of words and by reproducing ideas in their own words;
2. to explain the non-literal meanings, connotations and associations of figurative language;
3. to explain the reasons why the writer has used different registers of language;
4. to recognise patterns used in the structure of the passage, e.g. statement – illustration, statement – contrast, opinion – reason, etc.;
5. to recognise and reproduce in summary form the logical sequence of ideas in the passage;
6. to recognise and reproduce the significant points of an argument, as opposed to digressive, illustrative or secondary ones;
7. to explain the reasons for particular usages of punctuation or grammatical forms, or for particular types of sentence structure;
8. to be able to use the context of the passage to find clues to the meaning or implication of various types of linguistic usage;
9. to be familiar with such linguistic concepts as are necessary to discuss the workings of language in specific contexts (noun, adjective, verb, subject, complement . . .).

Teaching Aims

The examination implies –

1. that pupils should be experienced in 'close reading': i.e. that they should be used to responding to every means the writer has used to communicate his meaning;
2. that they should have enough 'technical' concepts to be able to indicate what these various means are.

Second List of Skills

After teachers' reactions to these 'objectives' had been sought in an 'open' questionnaire, it was intended to try to survey the views on S4 'objectives' of a bigger sample, using a fixed-response questionnaire. The results of this second questionnaire are, like those of the first, reported in Chapter I.

The need to write it led to a more detailed list of skills than the one based solely on O-grade examinations. At this stage the work of several other researchers and teachers of English came to the aid of the research team. Some of the influences are obvious in reading the *Second List of Skills*, some were less direct. The researchers are aware that, as far as the *Second List* is concerned, they are indebted at least to the following:

James Britton and his colleagues at the London Institute of Education.

The writers and compilers of the material used in the Open University's Reading Development Course.

The Scottish Central Committee on English, for ideas set out in various 'Bulletins'.

The authors of 'Assessing Comprehension', a discussion pamphlet of the London Association for the Teaching of English (Blackie & Son, 1968).

F. P. Robinson, for ideas contained in *Effective Study* (Harper & Row, New York, 1961).

L. E. W. Smith, for ideas in *Towards a New English Curriculum* (Dent, 1972).
Anthony Adams and John Pearce, for ideas in *Every English Teacher* (Oxford University Press, 1974).

The following is a slightly adapted version of the questionnaire sent to the teachers. The actual questionnaire (with results) is printed in Appendix 1(b).

NOTE

Speaking, reading, writing, and the assessment of success in these activities, are complex functions involving sets of skills, personal characteristics and experience. Most pieces of writing, for instance, will show that pupils have exhibited several of the skills set out in these lists. The classification used will, it is hoped, clarify what 'performance' in English involves, but it would not be desirable, or possible, to assess the achievement of each skill separately. Further, the list is not comprehensive. Some factors affecting English 'performance', such as the stage of intellectual and emotional development a pupil has reached, cannot be accurately measured with existing testing tools. There are also, even at O-grade standard, qualities recognised as intrinsic to good communication – e.g. breadth and aptness of vocabulary and variety and appropriateness of sentence structure – for which specific criteria cannot be formulated in advance.

WRITING

Three main headings are proposed, taken from James Britton's 'What's the Use' in *Language and Education* (Routledge and Kegan Paul, for the Open University).

1. 'Transactional' writing: to direct, question, get things done, participate in society . . .
2. 'Expressive' writing: to exchange or reveal feelings and opinions, convey attitudes, reveal personality . . .
3. 'Poetic' writing: to construct linguistic artefacts – organised language – as a means of trying to explore and master the complexities of reality.

'TRANSACTIONAL' WRITING

Pupils might be expected to:

1. Write instructions for a simple task (recipe, care of sports equipment . . .).
2. Formulate rules for a club, social evening, or some situation in which group behaviour must be controlled.
3. Write notes on events as they happen.
4. Report events factually.
5. Write a report presenting information logically.
6. Order and paragraph ideas.
7. Write an argument based on facts for or against a point of view.
8. Present evidence and draw conclusions from it.
9. Give arguments for *and* against a point of view and draw rational conclusions.
10. Illustrate a general idea with a particular example.
11. Generalise from particular instances.
12. Develop an idea, illustrating, quoting, adducing evidence in support.
13. Write a logical, theoretical argument, given premisses.
14. Write advertising copy (e.g. for school events . . .).
15. Write a persuasive argument using emotive language to win support.
16. Write a speech for or against a motion.
17. Summarise factual information.
18. Summarise an argument, maintaining the logical thread.
19. Write personal letters of various kinds.
20. Write formal letters to achieve practical ends.
21. Use the appropriate register of language for audience and purpose.

'EXPRESSIVE' WRITING

Pupils should be able to:

22. Describe their own feelings about or stimulated by events they have experienced.
23. Define their attitudes to aspects of society which impinge upon their experience (e.g. authority, other groups, other races).
24. State opinions on controversial topics.
25. Show awareness of the causes of their feelings and attitudes.
26. Convey their interest in various topics by writing knowledgeably about them.
27. Write fantasy stories (in which their own feelings, wishes or desires are significant).
28. Write imaginatively (i.e. with originality) or critically, expressing a personal view, given an artistic or natural stimulus (film, poem, music, picture, physical experience . . .).

'POETIC' WRITING

Pupils should be able to:

29. Write stories with some exposition of the causes and effects of events.
30. Organise events in a story to create a climax.
31. Create suspense in a story.
32. Write stories in which the feelings of more than one character are revealed.
33. Write stories with genuine interplay between events and characters.
34. Reveal character in a story through dialogue.
35. Dramatise (not just describe) events in a story.
36. Write a story with a moral or message.
37. Write adopting the point of view of someone with feelings and attitudes different from their own.
38. Write in simple poetic forms (e.g. ballad, haiku, free verse . . .).
39. Write in stricter poetic forms (e.g. blank verse, sonnet . . .).
40. Exploit (in prose or verse) the emotive overtones of words.
41. Use language figuratively.
42. Exploit the sound and rhythms of language to make a stronger impression on the reader.

Knowledge of appropriate vocabulary and of a variety of sentence structures is inherent in the objectives listed above and the skill with which this knowledge is employed would be assessed in judging the standard of any piece of writing by comparing it with other work. 'Technical correctness' would also be taken into account. While it is not practical to specify in detail minimum criteria in grammar and spelling, the following list indicates some skills contributing to 'correctness'.

Pupils in their normal writing might be expected to:

43. Use full stops correctly in nearly every instance.
44. Use the question mark correctly when required.
45. Use quotation marks for Direct Speech.
46. Use the comma correctly to mark off subordinate clauses and parenthetical or appositional phrases.
47. Use the comma correctly in combination with quotation marks.
48. Use the exclamation mark correctly.
49. Use the colon correctly.
50. Use the semi-colon correctly.
51. Use the apostrophe correctly.
52. Use quotation marks correctly for slang, foreign phrases . . .
53. Use parentheses correctly.
54. Use the dash correctly.
55. Spell correctly.
56. Show evidence of command of various grammatical devices, such as tenses, subordinate clauses, participles, noun phrases, linking words, contributing to the complexity and variety of sentence structures.

GENERAL READING

As regards their reading, in or out of school, pupils might be expected to:

57. Reproduce the gist of a narrative.
58. State their own emotional reaction to the content of what they have read.
59. Recognise or state the writer's attitude to his subject, and some of his ideas.
60. Recognise or state the tone/feeling of what they have read.
61. State their emotional and/or moral reaction to the writer's treatment of his subject.
62. Recognise and/or comment on the accuracy with which literature reflects human characteristics, relationships and problems.
63. Recognise and/or refer to the writer's techniques (e.g. devices for creating suspense, revealing character, or structurally relating parts to the whole, his use of figurative language, or other means he uses to make an emotional and/or moral impression on the reader or to create an imaginative world.
64. Give reasons for enjoying what they have read in terms of the writer's skill.

65. Give reasons for the judgement that one piece of writing is superior to another.
66. Locate specific information in a book using Contents, Index, Chapter headings etc.
67. 'Skim' read to get a general impression of a passage.

COMPREHENSION

When carrying out 'close reading' of a piece of writing, pupils might be expected to:

68. Recognise or recall information explicitly stated.
69. Paraphrase or translate such information.
70. Make inferences based on information explicitly stated and on their own experience.
71. Recognise or state the main points presented, as opposed to digressive, illustrative or secondary ones.
72. Reorganise and paraphrase (i.e. summarise) the main ideas presented.
73. Make a judgment based on their own knowledge, experience and values about the validity of ideas in a test.
74. Distinguish factual writing from opinion or persuasion.
75. Recognise or state the effect on meaning or tone in context the use of different registers of language.
76. Recognise or state the effects in context of particular sentence structures.
77. Recognise or state the effects in context of particular usages of grammar and punctuation.
78. Explain the implications, overtones, connotations in context of language used figuratively.
79. Recognise or state the general idea represented by a particular incident, detail or image.

Pupils' knowledge of language might be expected to include the following concepts; i.e they might be expected to be able to recognise or explain the function of each in context, even if they are not familiar with the terminology:

80. Noun.
81. Verb.
82. Adjective.
83. Adverb.
84. Pronoun.
85. Sentence.
86. Clause.
87. Adverbial clause.
88. Adjectival clause.
89. Noun clause.
90. Phrase.
91. Subject.
92. Object/Complement.
93. Singular/Plural.
94. Tenses.
95. Past and Present Participles.
96. Finite/Infinite moods.
97. Subjunctive mood.
98. Active and Passive voices.

(*d*) *Draft Plan for a Criterion Test*

For assessment in the 'comprehensive testing procedure', pupils might have been asked to complete all the following tasks, which together would have tested all the items in the *Second List of Skills.*

Possible Criterion Test Tasks

WRITING

Pupils might be asked to:

(1) Write factual reports, organising material logically (Skills 1-6).
(2) Make a case for a view, present an argument (7-13).

(3) Write persuasively, using emotive language, rhetorical devices . . . (14-16; 40-42).

(4) Summarise facts and arguments (17-18).

(5) Write letters for various purposes (19-20).

(6) Write to recall events and atmosphere of past experiences (21).

(7) Write to communicate personal feelings, opinions, attitudes, fantasies (21-27).

(8) Write stories (26; 28-37).

(9) Dramatise (33-34; 36).

(10) Write poetry (39-42).

In all the above the following would be taken into consideration in an assessment:

Breadth and appropriateness of vocabulary.

Appropriate use of formal/informal, emotive or figurative language, and of the sound and rhythms of language (37; 40-42).

Organisation and paragraphing of material (6).

Correctness of spelling (55).

Correctness of grammar (56).

Correctness of punctuation (43-54).

Complexity, variety and appropriateness of sentence structures (56).

READING

As far as reading complete works is concerned, pupils might be asked to:

(1) Reproduce the gist of what they have read (57).

(2) State personal reactions to the content and to the writer's treatment of it (58-61; 65).

(3) Comment on characters, relationships, problems portrayed (62; 65).

(4) Comment on style/techniques (63-65).

They might also be asked to:

(5) Keep a record of all their reading in and out of school (e.g., monthly). Checks could be made to counteract cheating by regular random questioning of pupils on the books they have entered in the record.

As far as 'close reading' is concerned, pupils might be asked to:

(6) Recognise, recall, translate, explicitly stated information ('Literal comprehension') (68-69).

(7) Make inferences based on the text. ('Inferential Comprehension') (70; 74).

(8) Make judgments about the validity of ideas in the text. ('Evaluative Comprehension') (73).

(9) State the effects on meaning and tone of various aspects of the writer's skill (74-79; 80-98) ('Appreciative Comprehension').

(10) Summarise the main ideas in the text (71-72).

The first 'blueprint' for the Criterion Test was designed to include all these tasks, though some were grouped together in feasible test papers, e.g., as 'Factual' or 'Persuasive' writing. The Criterion Test at this stage was shaped like this:

First Draft Plan for Criterion Test

WRITING

(1) *Expressive* (Possible Criterion Test tasks Nos. (6) and (7).)	1 from 6 or 7 options. Picture and verbal stimuli. Completely 'open' rubric – 'relevance' to be interpreted very broadly. Credit to be given for revelation of personality, individuality, confidence, sincerity . . . (judged from the way the language is used). Likely content: personal experience, exploration of feelings, attitudes, their causes and effects. Might be in narrative, descriptive or reflective form.
(2) *Factual* (Tasks Nos. (1), (4), and (5).)	1 from 4 options. Probably verbal stimuli only, though picture possible. Credit to be given for relevance, organisation of material, clarity, appropriate lay-out and style for prospective audience and function. Short piece of writing. Might be in letter or report or magazine article form.
(3) *Persuasive* (Tasks Nos. (2), (3), (4).)	1 from 4 options. Picture and verbal stimuli. Credit to be given as in (2), and for ability to use evidence, impress or move by emotive use of language, illustrate . . . Short piece, possibly in form of school magazine article.
(4) *Poetic* (Tasks Nos. (8), (9), (10).)	Story/Fiction for literary effect. Unlimited time (e.g. over a two or three week period). Credit for: characterisation, structure, plot, suspense, etc. . . Might be incomplete when assessed. Stimulus to be the beginning of a story, written by the Research Team (4 options), though pupil would be free to make his own start if he wished.

READING

(1) *Response to Literature* (Tasks (1) - (5); involving also tasks (6) - (10).)	a. 'General' questions on Reading, similar to those set in O-grade Paper I. b. 3 texts provided by SCRE: one short story, one poem, one self-contained extract from a play. Questions on each designed to test at a simple level appreciation of these texts as literature – i.e. questions mid-way between very specific ones referring to particular words/phrases, and very general ones, which seem to encourage regurgitation of content. Credit, e.g. identification with characters, evaluation of a character's behaviour, personal emotional reaction, etc., (the criteria for credit depend on the specific material used).
(2) *Comprehension* (Tasks (6) - (10).)	3 or 4 passages with questions covering all comprehension skills, including summary. (Might also include a test of awareness of the way language is used in advertising.)

There was, in addition, the possibility of subjecting the pupils to a standardised vocabulary test and a verbal reasoning test. Formal tests of spelling, punctuation and grammar could also have been devised to complement the assessment of the pupils' writing.

It was this scheme for a dauntingly massive test which was matched against reality in the form of the testing material which had survived pre-tests and expert evaluation, the expected limit of pupils' endurance, the time schools would allow, the finance available, and the need for O-grade examination practice. To make it fit better, the following decisions were taken:

1. 'Poetic' writing would be dropped. The schools were unlikely to be willing to accept the imposition of a three- or four-week long exercise (except for the one school in the experiment where such an assignment was

normal). There would be difficulties in finding time and money for reliable marking. Pre-tests in other schools had produced a little very interesting work, but seemed to show that success in writing fiction requires a kind of teaching which was not prevalent, at least in those particular schools.

2. There would be no standardised tests or formal tests of 'correctness'. With limited time, it was thought that these had a low priority as testing something other than normal performance in school or examination contexts.
3. The number of options in the 'Factual' and 'Persuasive' writing tests would be reduced to two or three each, and verbal stimuli only would be used. These decisions were taken for two reasons: (1) in order to reduce the range of writing possible, so that pupils could not easily slip into 'expressive' writing through misinterpretation of the stimulus, and so that marking could be more easily standardised; (2) because test stimuli successfully pre-tested and/or approved by the consultants were in short supply.
4. The number of options in 'Expressive' writing was, by contrast, to be increased, to allow as much freedom as possible to pupils in that section of the Test. Picture stimuli were to be used for the same reason.
5. One story only would be included in the 'Close Reading' test described under 'Response to Literature – b'. The removal of poetry and drama from this test was to save time. A short story seemed the most appropriate text for a wide ability range.
6. For the comprehension tests, advantage would be taken of the Research Officer's involvement in the SCEEB sub-committee on multiple-choice testing for O-grade. Passages and multiple-choice questions validated in SCEEB pre-tests of material developed by the sub-committee would complement a 'traditional' interpretation paper, which would, however, be designed so that the second passage sought specifically to test awareness of *structure* in the writing.
7. The first day's papers of the two-day Criterion Test would follow the pattern of the O-grade examination, so that it provided some examination practice for pupils whose schools had not set 'prelim' examinations.

II. The Criterion Test

The Criterion Test finally consisted of five papers, lasting in all some 7½ hours. Its elements and the skills they were each intended to test are shown in the following table. The assignments and questions set and the marking criteria applied can be scrutinised in detail in Appendix 4, where the test papers and marking schemes are printed.

The papers were taken by the pupils in numerical order, but are grouped here under 'Writing' and 'Reading' so that the skills assessed can be considered in cohesive groups. It will be seen that the Criterion Test obliged the pupils to show the extent of their command of a much wider range of skills than any one pupil needs to call upon in sitting an O-grade examination. The Criterion Test was, however, not absolutely comprehensive and it still allowed some choices of task, so that pupils obtaining the same mark for a paper may each have been assessed on (at least some) different skills: this phenomenon is, of course, even more likely to occur in both the O-grade and any internal assessment scheme and contributes to the imprecision of discriminatory assessment in English.

The Criterion Test: Description of Papers Set and Skills Sampled

WRITING

Test element	*Skills* (numbers refer to *Second List of Skills*; see page 114)
PAPER I(A) Composition. 1 hour Stimuli=pictures, quotations, titles open to various interpretations, intended to allow as much freedom as possible to the pupil to write about something which he knows and cares about. Instruction to write a composition revealing thoughts, feelings, experience or opinions. 15 minutes allowed before start to make choice of topic.	(Some of) Describe own experiences, feelings (22). Define own attitudes (23). Express opinions (24). Explore causes of feelings and attitudes (25). Exhibit interest in and knowledge about topic chosen (26). Reveal individuality, imaginativeness (28), perhaps in fantasy stories (27), involving own feelings, wishes, desires. (Possibly) story-writing skills (29-37). Command of vocabulary (not itemised) and sentence structures (56). Selection and organisation of material (6). Technical correctness (43-55). (Possibly) exploitation of figurative and emotive language (40, 41).
PAPER III(A) Factual Writing. 30 minutes One from two options: a specifically defined writing task, with indication in the question of the 'purpose' and likely 'audience'.	(Some of) Write instructions for a task (1). State rules of sport or game (2). Report, presenting information logically (5). Summarise factual information (17). (Perhaps) illustrate general idea with particular example (10). Order and paragraph ideas (6). Use appropriate register (21). Command of vocabulary (not itemised) and sentence structures (56). Technical correctness (43-55).
Test element	*Skills*
PAPER III(B) Persuasive Writing. 30 minutes One from three options. Specific task, with 'audience' and 'purpose' indicated, the latter being to influence others' actions by effective arguments.	(Some of) Argument based on facts (7). Draw conclusions from evidence (8). Give arguments for *and* against and draw conclusions (9). Use emotive language, rhetorical devices to strengthen argument (15). (Perhaps) write effective advertising copy (14). Illustrate ideas with examples (10). (Perhaps) generalise from particular instances (11). (Perhaps) develop an idea, illustrating, quoting, adducing evidence (12). Order and paragraph ideas (6). Use appropriate register (21). Command of vocabulary (not itemised) and sentence structures (56). Technical correctness (43-55).
PAPER II Interpretation and Language. 1 hour, 35 minutes Passage 1 (from 'To Kill a Mockingbird' by Harper Lee)	(All of) Literal comprehension (68). Translation (69). Inferential comprehension (70). Awareness of implications of register or tone of language used (75).

Passage 2 ('The Excavation of Ur' by C. W. Ceram)	(All of) Literal comprehension (68). Translation (69). Inferential comprehension (70). Summary (71-72). Appreciation of effects of sentence structures (76). Appreciation of effects of usages of grammar and punctuation (77). There was also concern to test awareness of means used by the writer to relate paragraphs to one another structurally (63).
PAPER IV Comprehension (Multiple-choice). 1 hour, 15 minutes	
Passage 1 ('Fake Antiques')	(All of) Literal Comprehension (68). Inferential Comprehension (70). Recognise main points (71). Recognise implications for meaning of register of language used (75).
Passage 2 ('Natasha')	As Passage 1, plus Recognise tone (60) Recognise writer's techniques for revealing character and creating tone (63). Recognise writer's attitude to subject (59).
Passage 3 ('Television')	As Passage 1.

Test element	*Skills*
PAPER V 'Close Reading' of Short Story (unlimited time) The story was read aloud to the pupils, who each had a copy of it. The instructions asked them (1) to re-read it with certain purposes in mind; (2) to answer 'in their heads' some questions designed to help their understanding of it; (3) they were then allowed (within reason) unlimited time to write answers to the assessment questions, which had been derived directly from the purposes suggested in (1) above.	(All of) State own emotional and moral reaction to the story (61). Appreciate and comment on the feelings/problems of the characters (62). Recognise the writer's attitude to the characters and events in the story and show that it influenced own reactions to them (59). Recognise or state techniques used by the writer to reveal character, hint at social background, prepare reader for the climax of the story . . . (63).
PAPER I(B) General Reading Questions. 1 hour, 15 minutes 3 questions, one from each of 3 sections, Prose, Poetry, Drama, each containing 3 options. 'Open' questions, no texts specified – similar to O-grade examination questions.	(Some of) Reproduce gist of a narrative (57). State personal reactions to content (58) and to writer's treatment of it (61). Recognise emotions in fictional characters and show understanding of relationships between them (62). Recognise and refer to some aspects of the writer's style/techniques (63). Appreciate tone, mood, atmosphere (60) (in Section III, Drama, question 1).

Skills not covered by the Criterion Test

Items included in the *Second List of Skills* but excluded from the Criterion Test were:

WRITING: *Skills not tested*

3. Write notes on events as they happen.
13. Theoretical argument.
16. Speeches.
18. Summary of someone else's logical argument.

19 & 20. Letters (most relevant skills were tested in other tasks).

38 & 39. Write poetry.

Though some pupils might have attempted some of the following skills in the Criterion Test (and some did), it was not intended that the Criterion Test *require* pupils to show them.

Skills usable in the Test but not required

9. Argue for *and* against a view and draw rational conclusions.
10. Generalise from particular instances.
11. Develop an idea in a sophisticated way.

29 - 37. Write successful literary stories.

40 - 42. Exploit emotive, figurative, rhetorical qualities of language.

READING: *Skills not tested*

65. Give reasons for the judgment that one piece of writing is superior to another.
66. 'Survey' skills – finding information in a book.
67. 'Skim' reading.
73. Make judgments about the validity of ideas in a text (evaluative comprehension).
74. Distinguish factual writing from opinion or persuasion.

80 - 98. Show specific knowledge of various grammatical concepts.

Skills usable in the test but not required

59 & 63. Comments on writer's attitude, style and technique were required, but at a fairly low level of sophistication.

64. Give reasons for enjoying what has been read in terms of the writer's skill.
78. Comment on figurative language.
79. State general idea represented by particular image, incident or detail.

III. The Marking of the Criterion Test

(1) *Double-marking*

The researchers began with the intention of adopting multiple-marking, on the recommendation of 'Schools Council Bulletin No. 12, Multiple-marking of English Compositions', by James Britton, N. Martin and H. Rosen. The cost of triple-marking having been found greater than the project could afford, double-marking was decided on, its clear superiority over single-marking being confirmed by a study conducted by Robert Wood and B. Quinn of the University of London School Examinations Board[1]. There was no marker-unreliability in Paper IV (multiple-choice comprehension test). For Paper II, Interpretation and Language, the most experienced markers were chosen to carry out single-marking, the marking team of eight containing three of the Principal Examiner's team and the two research officers.

(2) *Personnel*

A team of teachers was selected by the researchers from lists provided by SCEEB of those with good or, at least, satisfactory records as O-grade markers. Three criteria of selection were employed: (1) that the marker should have marked the relevant O-grade Paper satisfactorily for at least two years; (2) that he should not teach in one of the project schools; (3) that he live within fairly easy travelling distance of Glasgow or Edinburgh. Most of those employed were regarded by the SCEEB examiners as fully consistent markers; a few were only adequately consistent; and a number tended to severity or leniency, but could have their marks adjusted by marker-standardisation. The majority had several years' experience as markers. Three current members of the Principal Examiner's standardising team were in the group, which could be said to contain most of the best O-grade markers resident in Strathclyde, the Lothians, Central Scotland and Fife, with one outlier in Aberdeen. In all, 36 teachers were involved, including two of the researchers who were experienced examination markers.

(3) *Division of Labour*

Papers in the Criterion Test were assigned to markers as follows.

Paper I (A and B, Composition and Reading)
First marking: 8 markers
Second marking: 9 markers

Paper II (Interpretation and Language)
One marking only: 8 markers (including 3 examiners and the 2 researchers)

Paper III (Factual and Persuasive Writing) and V (Short Story) together
First marking: 8 markers
Second marking: 8 markers

Paper IV (Comprehension, multiple-choice) – marked by computer

(4) *Procedure*

Since the Criterion Test was intended to perform the same function as the O-grade examination and test some other skills as well, the marking procedure was kept as close as possible to that employed by the SCEEB. Marking instructions were written for each paper by the setters (i.e., the researchers) and these were distributed to markers, with photocopied sample scripts, in advance of markers' meetings. These meetings were, however, decision-taking, unlike the

---

[1] Wood and Quinn (1976).

SCEEB markers' meetings: i.e., each group of SCRE markers set its own standards for the sample scripts and, in the case of Paper II (Interpretation and Language), the marking instructions were modified as a result of the group's discussions. The Second Marking took place some three months after the First (the O-grade having intervened). One or two markers and the researchers, who conducted the meetings, were involved in both First and Second Markings. Almost all marked the 1977 O-grade examination. The influence of these factors may have been the reason why there was very little difference in the standards set by First and Second Markers on the sample scripts. For the Paper I sample scripts, the First Marking mean score was 30.7/60; the Second Marking mean score was 31.1/60. For the Paper III sample scripts, the First Marking mean score was 11.6/20; the Second Marking mean score was 11.5/20. For the Paper V sample scripts, the First Marking mean score was 10.5/20; the Second Marking mean score was 10.6/20.

This apparent unanimity of standard on sample scripts was not, however, maintained between First and Second Marking over all the candidates: the matter is discussed in Chapter VII.

Appendix 4
Criterion Test Papers and Marking Schemes

SCEEB/SCRE O-grade English Project
'Criterion Test'
PAPER I
A - Composition - 1 hour

INSTRUCTIONS In the following pages you will find some quotations, essay titles and pictures.

Write **ONE** composition showing what you have thought and felt about a quotation or title or picture or, for instance, a quotation and a picture together.

Your composition might be a story, a description of your own experience and your feelings, a statement of your views, or anything which will show what you have thought. (Do not simply describe what is in a picture or what a quotation says.)

You have 15 minutes to make your choice before the test begins.

No. 1

'I am alone with the beating of my heart.'

No. 2

'Behold, this dreamer cometh.'

No. 3

'I remember, I remember . . . '

No. 4

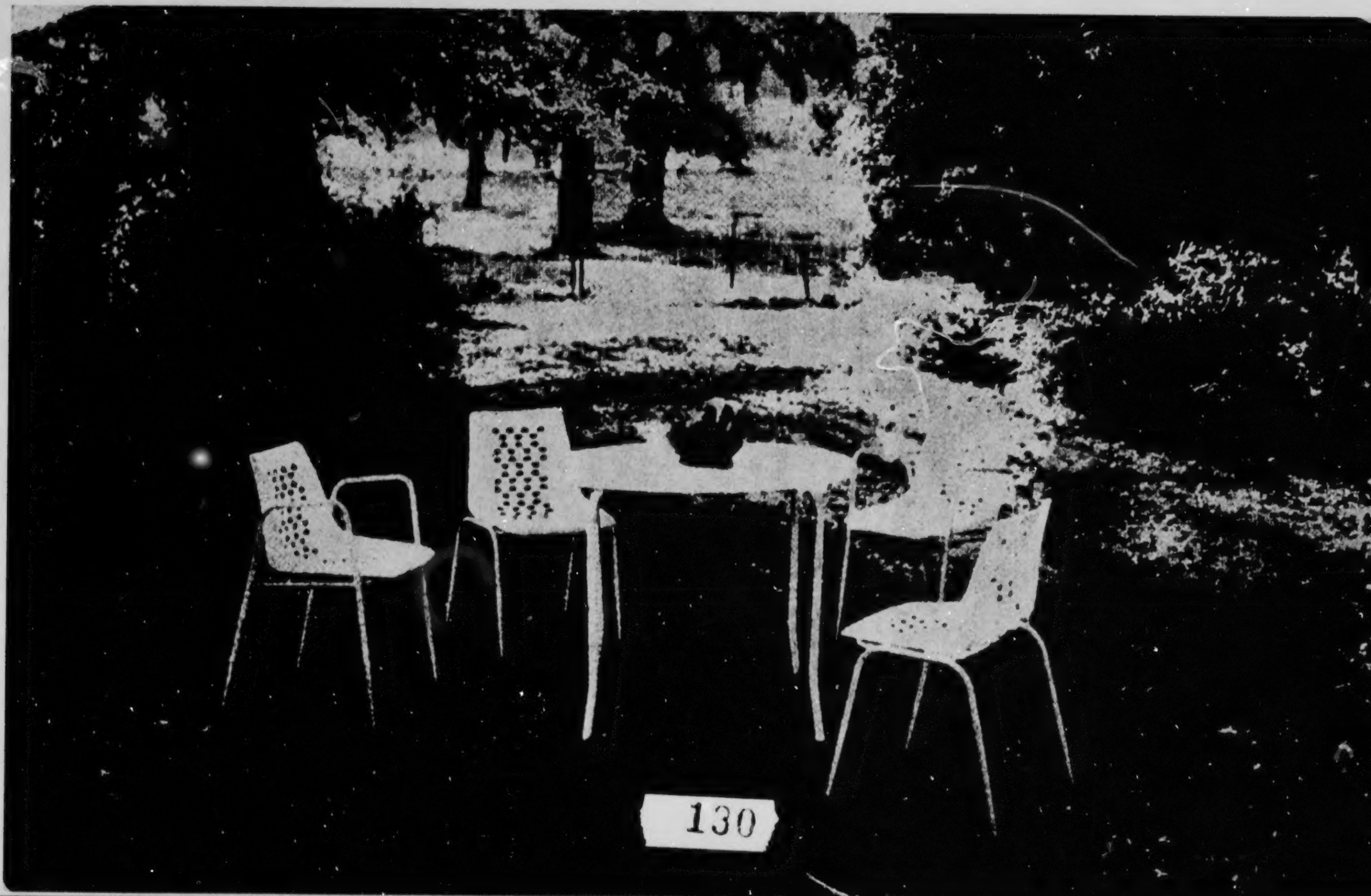
'Once I am sure there's nothing going on
I step inside, letting the door thud shut.
Another church: matting, seats, and stone,
And little books; sprawlings of flowers, cut
For Sunday, brownish now; some brass and stuff
Up at the holy end; the small neat organ;
And a tense, musty, unignorable silence,
Brewed God knows how long. Hatless, I take off
My cycle-clips in awkward reverence.'

No. 5

Threat.

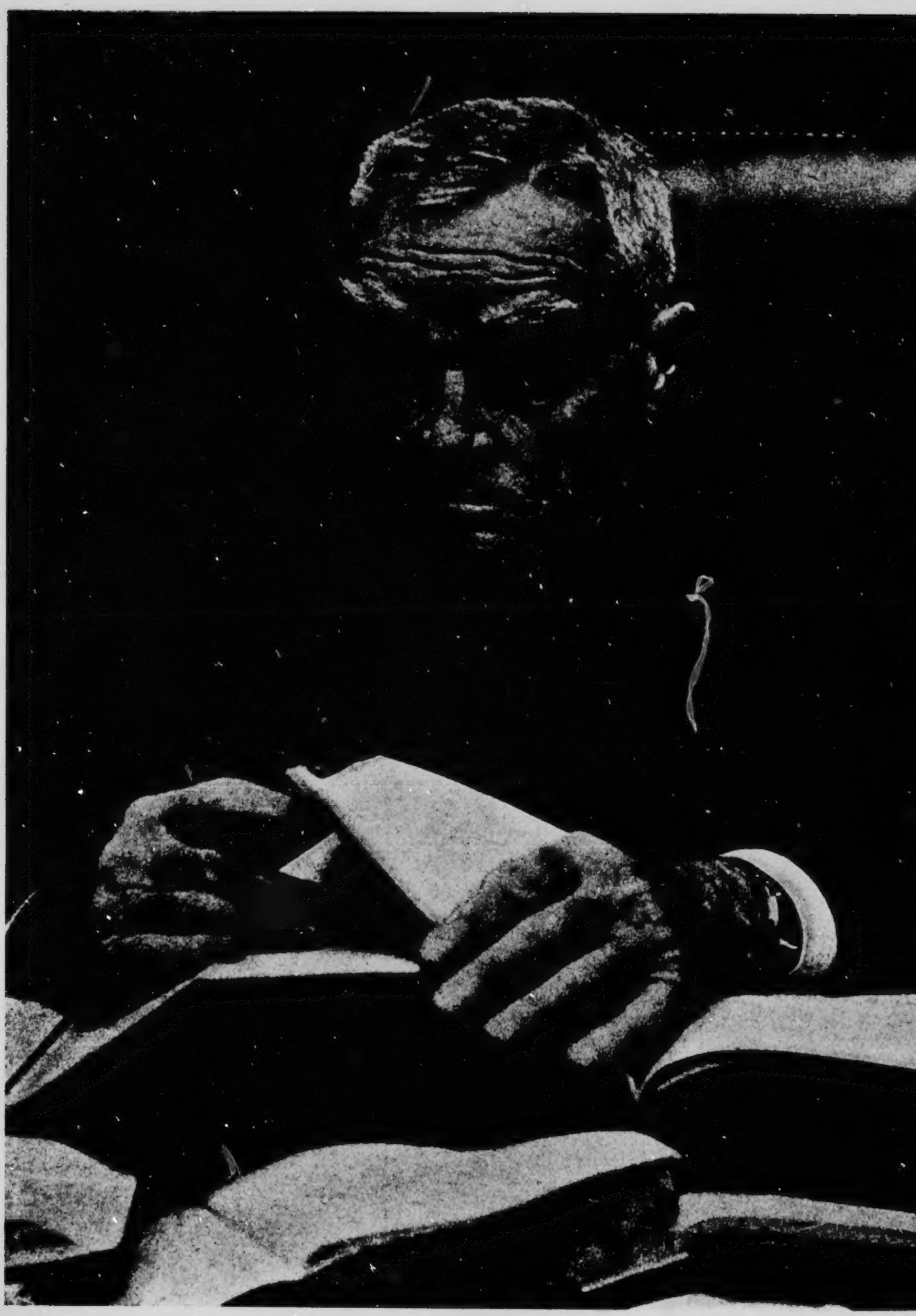
No. 6

Getting away from it all.



130

Open Window



134

Demo

SCEEB/SCRE O-grade English Project
'Criterion Test'

PAPER I

B - Reading - 1 hour 15 minutes

INSTRUCTIONS Answer **THREE** questions, one from each section.

Note: In each answer you should give the title and author of the book, story, poem or play you are writing about.

Section I: Prose

Answer **ONE** of the following:

(1) Show how a story or a novel you have enjoyed succeeded in gripping your attention.

Your answer should **briefly** summarise the story and also show why you found it interesting or exciting.

(2) Name a character in a story or novel whom you have felt you really understood.

Briefly describe the part played in the story by the person you have chosen. Then explain why the character appealed to you.

(3) Describe briefly the contents of a non-fiction book you have found interesting or useful. Then explain what features of the book specially contributed to its interest or usefulness.

You might discuss the way the material was presented as well as the content itself.

Section II: Poetry

Answer **ONE** of the following:

(1) Choose a poem which has excited or amused or disturbed you. Say briefly what it is about and explain as clearly as you can how it made its effect on you.

You might refer to the ideas in the poem, describe the way they are presented, mention any unusual or specially appropriate words, memorable images, etc.

(2) Choose two poems you have read on similar subjects. Say what each is about and what you like best about each. Point out any differences in the ways in which the poems present their subjects.

You may choose any subject you like for your two poems, but the following list may give you some ideas:

birds or animals;

machines;

human qualities, such as skill, bravery, strength, endurance, etc.;
feelings, such as fear, anger, hatred, love, envy, etc.

(3) Choose a poem which tells a story – it may be written as a ballad or in another form. Give a **brief** outline of the story and then explain how the reader's interest is held.

You might consider how suspense is created, description of character, imagery, incidents, etc.

Section III: Drama

Answer **ONE** of the following:

(1) Briefly describe a scene in a play which you have read and would like to see on stage. Then say how you think the scene should be presented so that the audience will enjoy it to the full; you should write your remarks under at least **two** of the following headings:

Movement:	how should individual characters move on stage? In what way should they come on or go off at any particular points in the scene?
Costume:	what clothes should individual characters wear to emphasise their personalities?
Lighting:	how should the scene be lit to help give the right mood?
Speech:	how should individual characters speak? Give your reasons.
Setting:	briefly describe a set which would help to create the right mood for the scene.

(2) Choose **two** characters who come into conflict in a play or television serial you know.

Explain briefly how the conflict comes about. Then show why it is almost bound to happen, considering the personalities of the people involved.

(3) From a play you have read or seen performed, select a character who either suffers misfortune or experiences happiness. Say how far you think the misfortune or happiness was deserved.

You should describe **briefly** what happens to the character in the play and then show what kind of person he or she is.

SCRE/SCEEB 'CRITERION TEST'

PAPER I

Guidance for Markers

General Points

1. *Please do not write anything on the scripts, or make any correction marks on them: simply record your scores for each pupil on the form provided.*

2. All composition and reading answers will be marked twice: it is enough to give an impression mark without *too* detailed attention to *specific* qualities or faults in the answers, though the guidance given below should be kept in mind.

A. COMPOSITION – 30 MARKS

I. APPROACH TO MARKING. The marker's approach to a composition should be positive rather than negative. His first task is to assess the positive merits of the piece of writing. *Only after due consideration has been given to the merits of the ideas and style should the marker assess the extent to which the errors revealed during reading damage the writer's power to communicate clearly and forcibly.* The final mark will therefore be an assessment of the *positive* merits of the composition *modified* by an assessment of the damage done by errors of grammar, spelling and punctuation.

In an attempt to give some guidance, a suggested grouping of marks into categories is given in Paragraph IV. This divides the range of marks into six broad categories and suggests the features which should characterise the compositions falling within each of the categories. The categories are broad ones: the precise mark given to any essay within a category will depend on the marker's assessment of the degree to which it meets the requirements of the category.

It must be stressed, however, that these categories are only generalisations for guidance which take no cognisance of errors in punctuation, spelling and grammar. When, for example, the quality and relevance of the ideas and handling of language would put a particular composition in one category, but the presence of faults of grammar, punctuation and spelling makes the marker feel an adjustment putting it in a lower category is necessary, he will obviously make such an adjustment. *Such adjustments should normally not exceed three marks.*

Where a composition is so very weak in some important respect that its writer's ability to communicate is impeded, the positive approach may have to be abandoned. Such a composition will not merit a pass whatever its good features may be. For example, a consistent failure to write in sentence form should be treated as a particularly damaging defect which cannot be redeemed by any of the other features. In such cases the final mark will be primarily a reflection of this defect. Such exceptional treatment, however, should be reserved only for compositions where *gross errors in expression seriously impede communication.*

No attempt should be made to allocate specific numbers of marks for specific aspects of the composition. Such a system often leads to a grand total quite out of proportion to the value of the work as a whole.

The range of marks should be wider than it frequently is in some schools. Generally, the best pupils should be expected to make high and occasionally full marks.

Markers who normally mark on a system where 10 or 20 is the maximum should bear in mind the need to make the mental adjustment to a system where the maximum is 30.

II. LENGTH. No precise length of composition is specifically demanded in the paper. The actual number of pages will, of course, depend on the size of the candidate's handwriting, but between two and three sides can be taken as an acceptable length, assuming average size of handwriting. Compositions which fall significantly below this should be penalised in proportion to the extent they fall short. *In general, such a penalty should not bring the mark below the pass line if the composition is worthy of a pass otherwise*. Compositions which are ludicrously short should, of course, receive a fail mark.

III. NOTE ON 'RELEVANCE'. The stimuli set in the test were chosen to give the pupils a very wide choice of topic and type of writing, with a view to allowing them to write as freely and 'expressively' as they wished. The relevance of the writing to the stimulus chosen should not therefore be a significant factor in the assessment in this part of the test. Markers should, however, consider 'internal relevance': their judgment of the quality of a composition should have taken into account the appropriateness of the content of the composition to *its own* topic or theme; they might, for instance, penalise a composition which lacked a unifying theme and consisted of separate ideas, observations or events written down without narrative or logical connection between them.

IV. SUGGESTED CATEGORIES FOR COMPOSITION WRITING. Two sets of categories are offered: the first is that provided by SCEEB for O-grade markers; the second is a less detailed categorisation which defines the qualities of scripts a little differently, though it does subsume the SCEEB instructions. Markers may use either or both. In any case, we should be glad to have the views of the markers on the relative merits of each categorisation.

SUGGESTED CATEGORIES FOR COMPOSITION WRITING (SCEEB). In the following categories the characteristics of each have been listed under the two headings, 'Content' and 'Form and Style'.

Under 'Content' the marker is asked to assess the quality of the ideas and their relevance to the chosen topic. Under 'Form and Style' the marker is asked to assess the positive contribution made by such things as the choice of vocabulary, the choice of sentence structure, and the ordering and paragraphing of ideas, to the effective expression of the content.

In practice, it is almost impossible to separate these two headings when marking an essay, and markers should never, under any circumstances, award separate allocations of marks to them. It is difficult to present any level of concept convincingly without a corresponding level of stylistic ability. Thus in composition marking it is normally found that the level of a candidate's performance under one heading is matched by the level of his skill under the other.

Markers should note that the heading 'Form and Style' refers to the assessment of the positive merits of expression. Errors in grammar, punctuation and spelling should be dealt with separately as outlined in Paragraph I.

CATEGORY I (27-30) – FIRST CLASS COMPOSITION WITH REAL SPARKLE

Compositions in this category will be distinguished by a sparkle in the handling of the topic which reveals something of the personality and forcefulness of the writer. Note, if a well written composition reveals as much maturity and forcefulness as can be reasonably expected from an O-grade candidate, there is no reason why it should not be awarded full marks.

Content Compositions will cover chosen topic very fully and show sincere interest in and knowledge of the subject. The point of the topic will be clearly grasped and all ideas will be relevant to it. In a story there will be a gripping and interesting development of character and situation.

Form and Style Compositions will be clearly paragraphed, with logical sequence of ideas. The sentence structure will express the ideas forcefully and the choice of vocabulary will be wide and apt. The composition will reveal a real and forceful command of language.

CATEGORY II (23-26) – VERY COMPETENT COMPOSITION WITH VERY FULL DEVELOPMENT OF SUBJECT

Compositions in this category will show a development of the chosen topic which is both full and relevant. However, the content and style will lack those recurring touches of forcefulness and personality which mark out a Category I composition.

Content Ideas will all be relevant to the point of the topic and of a high standard. However, they will not quite have the depth or sparkle of those in Category I. They will be sufficient in number to give a very full treatment of the topic. In the story there will be a very competent development of character and situation.

Form and Style The composition will be clearly paragraphed, with logical sequence of ideas. The command of suitable sentence structures and choice of apt vocabulary will reveal a very competent command of language.

CATEGORY III (19-22) – COMPOSITION WITH A REASONABLE DEVELOPMENT OF SUBJECT

Compositions in this category will show a reasonable number of relevant ideas. However, these will lack the range and fullness of Category II, and the sparkle and forcefulness of Category I compositions.

Content There will be a sufficient number of relevant ideas to give a reasonable development of the topic. The ideas will be more pedestrian than in the first two categories. In a story the development of character and situation will be reasonable but lacking the liveliness, freshness and fullness of Categories I and II compositions.

143

Form and Style Paragraphing will be reasonable, with the line of development easy to follow. The sentence structure will be sound, though perhaps lacking in variety and so in forcefulness. The vocabulary will be adequate, but overall revealing a less sure command of language than in the first two categories.

CATEGORY IV (15-18) – PASS GRADE COMPOSITION

Compositions in this category will show the lowest acceptable standard in the development of subject and in the handling of language.

Content There will be just enough relevant ideas to meet the demands of a pass grade, but their range and depth will leave much to be desired.

Form and Style The sentence structure and vocabulary will reveal only a basic level of literacy, i.e., an ability to communicate without the flow of ideas being *seriously* impeded by faults and infelicities in expression. Apart from this basic requirement, the use of language will reveal few other positive merits.

CATEGORY V (10-14) – COMPOSITION WITH MAJOR DEFECTS BRINGING IT BELOW PASS LINE

Content Ideas will fail to open up or cover topic adequately. There will be a failure to say anything of significance on chosen subject.

Form and Style The choice of sentence structures and vocabulary will be so limited that the writer has difficulty in expressing his ideas clearly and easily.

CATEGORY VI (9 DOWNWARDS) – OUTRIGHT FAILURE

Content Complete failure to get to grips with the subject **OR** a hopelessly short composition.

Form and Style Faults in expression will impede communication so seriously that there is a complete breakdown in the flow of the composition.

SUGGESTED CATEGORIES FOR COMPOSITION WRITING (SCRE)

VERY GOOD WORK	AVERAGE WORK	POOR WORK
Marks 30 —— 27, 26 —— 23, 22 ——	— 19, 18 —— 15, 14 ——	—— 10, 9 —— 0
● Many ideas, observations, of some maturity and depth, conering in a unified whole. Apparent personal involvement in the subject matter.	● Reasonable number of ideas, observations, perhaps indicating a lack of real knowledge of the subject, or seeming pedestrian, 'half-baked' or very derivative, though appropriate for the topic.	● Few ideas expressed and of little significance. Very little understanding of or interest in the subject shown.
● 'Authorship': involving (1) awareness of the reader on the writer's part and appropriate tone vis a-vis the reader. (2) Confidence of the writer in the value of his writing. (3) A 'distancing' of the writer from his experience, a greater explicitness, the creation of a context for events and feelings. (4) An orderly and selective structuring of the writing. (5) Individualised personalised use of language, involving broad and apt vocabulary and variety of appropriate sentence structures.	● 'Towards authorship': (1) Some attempts to write 'for an audience', though not sustained, or not quite properly pitched in tone. (2) A willingness to express oneself, but a sense of conviction is lacking in the writing. (3) Use made in the writing of personal experience or flights of imagination; attempts to set these in contexts, but not fully successful; some, but not much success in relating personal feelings or experience to the larger world. (4) adequate paragraphing: structuring often by chronological sequence. (5) Adequate vocabulary and sentence structures, but lacking absolute appropriateness, range and variety.	● No sense of 'audience' or structure; context inexplicit – just the ebb and flow of a few thoughts. Limited vocabulary and monotonous simple sentence structures.
● Correctness in nearly every instance (Spelling, Punctuation and Grammar).	● Generally correct Spelling and Grammar, correct use of full stops nearly always, perhaps with some faults otherwise.	● Many technical faults causing the flow of expression to be impeded, (e.g. complete or almost complete failure to use full stops correctly).

PAPER I: B

READING

Three questions, each to receive *an impression mark out of 10.*
Marking should take account of the following:

1. the presence or absence of evidence that the text has been read and understood;

2. the presence or absence of some *judgment* of the text, some emotive and/or intellectual and/or moral response to it;

3. relevance of the answer to the question set.

Notes: 1. Pupils should not be penalised for failing to use the approaches *suggested* in some questions, so long as they do answer both parts of the main question.

Questions suggesting, as opposed to requiring, certain approaches are: I 3, II 1, II 3.

2. Answers showing as much awareness of the writer's communication and techniques and as much maturity of judgment as can reasonably be expected of intelligent O-grade candidates should receive full marks.

A bare pass mark should be given when the answer shows adequate knowledge of the main events of the story and includes some sensible response to the second requirement of the question, without indicating that the pupil has reacted to the less obvious qualities of the text.

Answers which merely 'summarise the story', no matter how brilliantly, cannot score more than 4½/10.

3. The following points should be noted about particular questions:

I 3: *Non-fiction* book required.
If fiction is used, score out of 5.
Note that 'features' in the question (a) is plural, but (b) does not *necessarily* mean 'techniques' or 'style' or 'manner of presentation': a good answer to this question could be solely concerned with content.

II 2: *Two* poems must be discussed for a good mark.
This is quite a difficult question: markers should be fairly generous in their assessment of answers to the second part of the question.

III 1: A *scene* only is wanted, not a whole play (though attempts to deal with a whole play will probably be self-penalising). A hard question, since not much is done along its lines. Markers should be fairly generous in assessing responses to the second part, but *should* expect several sensible comments for a 'pass' answer: it should not be an 'easy option'.

III 2: Characters chosen should, of course, actually *be* in conflict, though this may be interpreted fairly widely, to include, e.g., 'having widely differing views on the same subject': they do not actually have to come to blows.

SCEEB/SCRE O-grade English Project

'Criterion Test'

PAPER II

Interpretation and Language - 1 hour 35 minutes

INSTRUCTIONS Read each passage carefully and answer the questions in your own words as far as possible.

PASSAGE A

The Return of Dill

(Scout is becoming more and more exasperated with her brother Jem's attitude to her . . . .)

His maddening superiority was unbearable these days. He didn't want to do anything but read and go off by himself. Still, everything he read he passed along to me, but with this difference: formerly, because he thought I'd like it; now, for my edification and instruction.

'Jee-crawling-hova, Jem! Who do you think you are?'

'Now I mean it, Scout, you antagonize Aunty and I'll – I'll spank you.'

With that, I was gone. 'You damn morphodite, I'll kill you!' He was sitting on the bed, and it was easy to grab his front hair and land one on his mouth. He slapped me and I tried another left, but a punch in the stomach sent me sprawling on the floor. It nearly knocked the breath out of me, but it didn't matter because I knew he was fighting, he was fighting me back. We were still equals.

'Ain't so high and mighty now, are you!' I screamed, sailing in again. He was still on the bed and I couldn't get a firm stance, so I threw myself at him as hard as I could, hitting, pulling, pinching, gouging. What had begun as a fistfight became a brawl. We were still struggling when Atticus separated us.

'That's all,' he said. 'Both of you go to bed right now.'

'Taah!' I said at Jem. He was being sent to bed at my bedtime.

'Who started it?' asked Atticus, in resignation.

'Jem did. He was tryin' to tell me what to do. I don't have to mind *him* now, do I?'

Atticus smiled. 'Let's leave it at this: you mind Jem whenever he can make you. Fair enough?'

Aunt Alexandra was present but silent, and when she went down the hall with Atticus we heard her say, ' . . . just one of the things I've been telling you about,' a phrase that united us again.

Ours were adjoining rooms; as I shut the door between them Jem said, 'Night, Scout.'

'Night,' I murmured, picking my way across the room to turn on the light. As I passed the bed I stepped on something warm, resilient, and rather smooth. It was not quite like hard rubber, and I had the sensation that it was alive. I also heard it move.

I switched on the light and looked at the floor by the bed. Whatever I had stepped on was gone. I tapped on Jem's door.

'What,' he said.

'How does a snake feel?'

'Sort of rough. Cold. Dusty. Why?'

'I think there's one under my bed. Can you come look?'

‘Are you bein’ funny?’ Jem opened the door. He was in his pyjama bottoms. I noticed not without satisfaction that the mark of my knuckles was still on his mouth. When he saw I meant what I said, he said: ‘If you think I’m gonna put my face down to a snake you’ve got another think comin’. Hold on a minute.’

He went to the kitchen and fetched the broom. ‘You better get up on the bed,’ he said.

‘You reckon it’s really one?’ I asked. This was an occasion. Our houses had no cellars; they were built on stone blocks a few feet above the ground, and the entry of reptiles was not unknown but was not commonplace. Miss Rachel Haverford’s excuse for a glass of neat whisky every morning was that she never got over the fright of finding a rattler coiled in her bedroom closet, on her washing, when she went to hang up her nightdress.

Jem made a tentative swipe under the bed. I looked over the foot to see if a snake would come out. None did. Jem made a deeper swipe.

‘Do snakes grunt?’

‘It ain’t a snake,’ Jem said. ‘It’s somebody.’

Suddenly a filthy brown package shot from under the bed. Jem raised the broom and missed Dill’s head by an inch when it appeared.

‘God Almighty.’ Jem’s voice was reverent.

We watched Dill emerge by degrees. He was a tight fit. He stood up and eased his shoulders, turned his feet in their ankle sockets, rubbed the back of his neck. His circulation restored, he said, ‘Hey.’

Jem petitioned God again. I was speechless.

‘I’m ’bout to perish,’ said Dill. ‘Got anything to eat?’

In a dream, I went to the kitchen. I brought him back some milk and half a pan of corn bread left over from supper. Dill devoured it, chewing with his front teeth, as was his custom.

I finally found my voice. ‘How’d you get here?’

By an involved route. Refreshed by food, Dill recited this tale: having been bound in chains and left to die in the basement (there were basements in Meridian) by his new father, who disliked him, and secretly kept alive on raw field peas by a passing farmer who heard his cries for help (the good man poked a bushel pod by pod through the ventilator), Dill worked himself free by pulling the chains from the wall. Still in wrist manacles, he wandered two miles out of Meridian where he discovered a small animal show and was immediately engaged to wash the camel. He travelled with the show all over Mississippi until his infallible sense of direction told him he was in Abbott County, Alabama, just across the river from Maycomb. He walked the rest of the way.

‘How’d you get here?’ asked Jem.

He had taken thirteen dollars from his mother’s purse, caught the nine o’clock from Meridian and got off at Maycomb Junction. He had walked ten or eleven of the fourteen miles to Maycomb, off the highway in the scrub bushes lest the authorities be seeking him, and had ridden the remainder of the way clinging to the backboard of a cotton wagon. He had been under the bed for two hours, he thought; he had heard us in the dining-room, and the clink of forks on plates nearly drove him crazy. He thought Jem and I would never go to bed; he had considered emerging and helping me beat Jem, as Jem had grown far taller, but he knew Mr Finch would break it up soon, so he thought it best to stay where he was. He was worn out, dirty beyond belief, and home.

You should use your own words, as far as possible, in answering the following questions:

1. (*a*) Explain the change in Jem's relationship with his sister which is referred to in the first paragraph.

(*b*) Why was Scout pleased that Jem fought back when she punched him? **(3)**

2. ' "Taah!" I said at Jem. He was being sent to bed at my bedtime.' (line 19)

(*a*) How does Scout feel towards Jem at this moment?

(*b*) Explain how you know this from the two sentences quoted. **(4)**

3. (*a*) What was Scout's usual attitude to her brother?

(*b*) How do you know this? **(4)**

4. (*a*) What was Jem and Scout's reaction to finding Dill under the bed?

(*b*) Why did they react as they did?

(Consider that Scout was 'speechless' and moved 'in a dream', and that Jem's voice was 'reverent' when he said 'God Almighty'.) **(4)**

5. Two accounts are given of how Dill had returned from Meridian to Maycomb. (lines 69-85)

(*a*) Why does Jem repeat Scout's question, 'How'd you get here?' ? **(1)**

(*b*) What was the real reason for Dill's return?
(Answer as fully as you can, considering Dill's relationship with Jem and Scout.) **(3)**

(*c*) Judging from the way Dill first tells the story, what kind of person would you say he is? **(3)**

6. What contrast is there between family relationships in Jem and Scout's family and in Dill's family? **(3)**

25

PASSAGE B

The Excavation of Ur

Then Woolley made his most sinister find: the Royal graves of Ur contained the remains of commoners as well as of royalty.

In one tomb lay a number of soldiers of the guard, wearing copper helmets and with spears in their bony hands. At the farther end of the chamber lay nine ladies of the court, still wearing the elaborate golden head-dresses that they must have donned for the funeral ceremony. By the entrance stood two heavy ox-drawn carts; in the carts were the drivers' bones, and at the oxen's heads lay the bones of the grooms.

In the grave of Queen Shub-ad ladies of the court were found lying in two parallel rows. At the end of one of these rows was a man's skeleton – that of the court harpist. His arm bones were still lying across his broken instrument, which was ornamented with a calf's head in lapis lazuli and gold. Apparently he had held fast to his instrument even as death overcame him. At the wooden bier where the Queen herself reposed, two female skeletons were found in a crouching position.

What did all this mean?

There was only one explanation: here the greatest possible sacrifice had been exacted of mortal men – their own lives. Woolley had stumbled on a scene of planned human sacrifice, carried out in conformity with the king-god principle. The position of the skeletons, as well as other circumstances of the find, indicated that the victims – courtfolk, soldiers, and servants – had died quite peacefully and it is thought probable that they walked to their places, took some kind of drug, and lay down; after the drug had worked, whether it produced sleep or death, the last touches were given to their bodies and earth was flung in and trampled down on the top of them.

What conclusions did Woolley draw from these finds? 'In no known text,' he writes, 'is there anything that hints at human sacrifice of this sort, nor had archaeology discovered any trace of such a custom or any survival of it in a later age; if, as I have suggested above, it is to be explained by the deification of the early kings, we can say that in the historic period even the greater gods demanded no such rite: its disappearance may be an argument for the high antiquity of the Ur graves.'

C. W. Ceram.

Use your own words as far as possible in answering the following questions:

1. (*a*) What was it that Woolley found 'sinister' about the graves at Ur? **(1)**

(*b*) Two things in particular about the graves required explanation. What were they? **(2)**

2. What explanation for the find is given in Paragraph 5? **(2)**

3. What belief made this practice acceptable to the people of Ur? **(1)**

4. According to the last paragraph, which two facts led Woolley to conclude that this burial place was exceptionally old? **(2)**

5. (*a*) Why is a colon (:) used in line 1? **(1)**

(*b*) Why is there a semi-colon (;) instead of a full stop in line 7? **(1)**

(*c*) '. . . the victims – courtfolk, soldiers and servants – had died quite peacefully . . . ' (lines 21-22). Explain why the dashes have been used here. **(1)**

6. (*a*) Why does the first sentence stand as a separate paragraph? **(1)**
   (*b*) Why is the question, 'What did all this mean?' also written as a separate paragraph? **(2)**
   (*c*) Explain why the last two paragraphs are separated from each other. **(1)**

7. Write a summary of the passage down to 'trampled down on top of them'.

   Do not include the last paragraph in your summary.

   The following plan may help in writing the summary:

   (1) What was found in the tombs at Ur.
   (2) The significance of the find: what it showed about royal burials there.

   Your final summary should be in one continuous paragraph of about 100 of your own words. **(10)**

25

SCRE/SCEEB 'CRITERION TEST'

PAPER II

Marking Scheme

PASSAGE A.

1. 2 marks for (*a*).
1 mark for (*b*).

(*a*) 2 marks for describing two *aspects of Jem's behaviour* which have changed **OR** for giving two *reasons* why he has changed, **OR** for giving one of each.

ASPECTS OF BEHAVIOUR: Any two of:

1. He no longer played with her (1) **OR** he stayed by himself (1) **OR** is isolated (1) **OR** reads a lot (1).
2. He treated her as an inferior (1) (was *superior* = ½) **OR** he treated her as a child (1) **OR** behaved as though he was her father or an adult (1).
3. He tried to teach her things (1) **OR** educate her (1) **OR** give her *instruction* (½).
   'passed on things for her edification and instruction' = 0.
   'bullied her' = 0.

REASONS FOR CHANGE:

1. He was growing up (1) **OR** felt adult (1) **OR** felt responsible (1) **OR** important (1) (superior = ½).
2. He was no longer interested in her (1). **(2)**

(*b*) It proved that they were still on the same level (1) **OR** that he wasn't really superior (1).

(equal or equals = ½ unless expanded – e.g., equal like two children in a fight = 1) **(1)**

2. 2 marks for (*a*).
2 marks for (*b*).

(*a*) Give **two** marks for clear statement of one of the following ideas: Triumphant (2) **OR** crowing (2) **OR** pleased at his humiliation (2) **OR** pleased that he's being taken down a peg (2) **OR** mocking (2).

Give **one** mark if the above ideas are absent but any of the following are present:

Hostile } not feeling friendly towards him (1) / not at one with him (1) **OR** spiteful (1)

OR trying to get at him, annoy him (1) **OR** pleased that he's made equal with her again (1) (without any implication that he has been *reduced* to her level).
Angry=0; annoyed=0; sarcastic=0.

(N B. For **two** marks, her sense of *triumph* or *mockery* must be noted.)

(2)

(*b*) One mark each for any two of the following:

1. An explanation that 'Taah!' is an expression of derision or of triumph (1) (sarcasm=0).
   'She sounds mocking or triumphant'=1, *if specifically linked to 'Taah!'*.
2. An explanation that 'at' has been used intead of 'to' (to convey antagonism) (1).
3. A comment on '*sent* to bed', showing that this humiliates Jem (1), e.g., 'he is sent to bed like a child'=1.
4. A comment that Jem is humiliated by going to bed at *a child's bedtime* (1) or at his *younger* sister's bedtime (1).

(Note: in 3 and 4 some awareness that it is *humiliating* for Jem must be conveyed.
e.g., 'I know it because he had to go to bed at her bedtime'=0.)

(2)

3. 2 for (*a*), 2 for (*b*).

(*a*) Any two of the following:

1. Trust (1) **OR** reliance for help (1) (or for protection (1)).
2. Friendly (1) **OR** e.g., 'they got on well' (1) (united=½).
3. She regarded him as an ally (1) **OR** e.g., 'they were on the same side' (1).

If (*a*) is wrong, give 0 for (*b*).

(2)

(*b*) *Answers to* (*b*) *must be related to answers to* (*a*).
Any two of the following:

1. She sought his help with the 'snake' (1).

2. Either reference to the friendly exchange of 'Good night' (1) **OR** reference to 'united us again'.
3. Reference to the fact that they were united *against Aunt Alexandra* in some sense (1).

N.B. Quotations are acceptable answers in (*b*).

(2)

4. 2 for (*a*), 2 for (*b*).

(*a*) Good answers should pick up not only the surprise of Jem and Scout but also the *intensity* of their amazement, the sense of *awe* or *unreality* which seemed to come over them. So:

Surprised=1; very surprised=1; amazed or astonished=1; absolutely astonished=2; couldn't believe their eyes=2; shocked=2; stunned=2; awestruck=2.
If (*a*) is wrong, give 0 for (*b*).

(2)

(*b*) One mark each for any 2 of the following:

1. They didn't expect him (1) **OR** they thought he was in Meridian (1).
2. The *place* in which he was found was unlikely (1) (e.g., 'they didn't expect to find anyone *under the bed*', **OR** they were expecting a snake, not Dill.
3. The situation was so unusual as to seem *unreal* (1).
4. They realised that there was a *seriousness* in Dill's presence there (1) **OR** that Dill must have had a very good reason for being there (1) **OR** that this was a situation which would be hard to explain (1) **OR** which could cause some trouble (1).

(2)

5. (*a*) He doesn't believe the first version (1). **(1)**

(*b*) 3 marks:

1. 1 is for picking up the very last phrase of the passage 'and home' e.g., 'Dill felt that Maycomb was his real home'=1; **OR** 'he used to live in Maycomb'=1.
2. The other 2 are for a gloss on 'home' – some explanation of what 'and home' implies. Give one mark for each of any two of the following:
   - Dill felt that Jem and Scout were like his brother and sister (1).
   - Dill felt at ease with Jem and Scout (1) **OR** happy with them (1) **OR** wanted to be with them (1).
   - he was unhappy in Meridian (1) (but **not** 'his father ill-treated him').
   - he trusted Jem and Scout (1) **OR** felt secure (1) with them.

- he felt they were concerned (1) **OR** interested in him (1) (**OR** interested *by* him (1)).
- he missed Maycomb (1) **OR** he missed Jem and Scout (1) **OR** he was very fond of them (1) (liked them=0).

(3)

Personal qualities/traits required, not just his situation.

(*c*) 3 marks: Any 3 of:

1. Imaginative (1) **OR** fantasiser (1) **OR** dreamer (1).
2. Story-teller (2) **OR** romancer (2) (this includes 'imaginative').
3. Liked to be the centre of attention (1) or liked to cause amazement (1).
4. Was really unhappy, but fictionalised it (1) or had poor relations with his father but fictionalised the fact (1) (**not** just 'had bad relationship with his father').
5. Had a sense of humour (1).

(3)

Liar=0, unless it is clear that the pupil means only 'as a fantasist' or 'for amusement' or 'as a way of coming to terms with unhappiness'.

6. Give an *impression mark* out of 3 (half marks allowed).
Points to be considered:

Dill's family:	1. Dill has poor relationship with his parents.
Jem and Scout's family:	2. Atticus displays understanding of Jem and Scout.
	3. Jem and Scout have basically a happy relationship.
	4. There is a mutual bond in Jem and Scout's family,
	5. despite surface disagreement.

Three marks for a convincing statement of *contrast* between the two families in terms of their internal relationships (not, e.g., their social standing or their money).
If no contrast=0.

(3)

25

PASSAGE B.

1. (*a*) One mark for the likelihood that the commoners had been put to death (1). Accept 'the fact that there were ordinary people there as well as royalty'=1.

(No penalty for 'lifting' the word 'commoners'.) **(1)**

(*b*) One mark for each of two points: one point about the *presence* of *ordinary* people (as well as kings and queens), and one point

about the *conditions or circumstances* in which the bodies were found.
Penalise lifts (½).

So:

1. **EITHER** – why were there ordinary people there? = 1
   OR why were so many buried in the same place? = 1
   OR why had the ordinary people died? = 1.

2. **EITHER** – a generalised reference to the *positions* of the bodies = 1, e.g., 'why were they lying in peaceful or natural positions?' = 1 **OR** 'why were they lying in the positions they were in?' = 1.

   OR a generalised reference to the formality of their *dress* = 1, e.g., 'why were they dressed so elaborately or richly?' = 1

   OR a generalised reference to the *objects* found there = 1, e.g., 'why had they been buried with the things they normally worked with, such as spears, carts and musical instruments?' = 1.

Particular examples = ½, e.g., 'why were they wearing head-dresses?' = ½ **OR** 'why was the harpist still holding his harp?' = ½ etc.

(2)

2. One mark each for any two of the following:

1. they had been *put to death* (or died) when the king or queen was buried = 1 (had been sacrificed = ½, it was a human sacrifice = ½). Must show understanding of 'sacrifice'.
2. they had died *willingly* (= 1) **OR** *calmly* (= 1) **OR** *without a struggle* (= 1) (peacefully = ½).
   as a matter of course = 1; suicide = 1.
   They had (willingly) sacrificed themselves when the king died = 2. (Accept 'sacrificed' if in the phrase 'sacrificed themselves'.) Must show understanding of 'sacrifice'.
3. because the king was a god (1) **OR** because the king needed his followers or servants after death (= 1) **OR** because they were ready to join the king in the next life = 1

(2)

3. They believed that the king was a god = 1 **OR** They believed in life after death = 1.
King-god = ½.

(1)

4. One mark for each of any two points:

1. There are no *written records* of the custom = 1

   OR it is not *recorded in history* = 1

   OR since history began no kings or gods have required human sacrifice = 1 ('in the historic period' = ½: extensive lifts to be severely penalised).

2. Archaeologists have never found any sign of the custom *at any later time* = 1 ('in a later age' = ½), or in any other place = 1.
   (Archaeologists have never found any sign of it = ½).
   No evidence in later times = 1.

(2)

5. (*a*) One mark for an answer which indicates awareness that the second part of the sentence is an *explanation* of the first part. E.g., the colon is used because the second part explains why it was sinister = 1, **OR** the second part states what Woolley found = 1.

(1)

(*b*) One mark for an answer indicating awareness that the second part is an *expansion of* or *elaboration on* the first part. E.g., the semi-colon is used because the second part (or the second sentence) goes on to say more about the carts = 1, **OR** '. . . is still referring to the carts' = 1, **OR** '. . . tells us more about the carts' = 1.

(1)

(*c*) One mark for an answer indicating awareness that the dashes enclose a parenthesis.
E.g., they separate the words from the rest of the sentence = 1.

OR the sentence would make sense without the words between the dashes = 1.

OR they are used to allow the writer to explain who the victims were without starting a separate sentence = 1.

Must show awareness of idea of parenthesis.

(1)

6. (*a*) One mark for **EITHER**

it is an *introductory* statement = 1

OR it states what the passage is going to be about = 1

OR it is a *general* statement, before the other paragraphs discuss the details of the find = 1.

(1)

(*b*) Two marks: one for awareness that, in a sense, the question sums up the preceding paragraphs, or indicates the *end* of a section; one for awareness that it introduces the answer.
So:

It is a *turning point* in the passage = **2** **(2)**

OR It *links* the two paragraphs above and below it = 2

It gathers together the writer's thoughts about what was found = 1, and it leads on to his explanation of the find = 1.

(*c*) Different topics = 1. One is explanation of the meaning of the find, one gives Woolley's conclusions about it = 1. **(1)**

7. SUMMARY

The summary should be given an impression mark out of 10, taking into account the presence or absence of the points listed below, fluency, linking words and lifts from the passage.

(10)

Main points:

Lines 1-15 . The presence of ordinary people, soldiers, courtiers and servants, in the royal tombs.
. The type of objects they had with them, and the dress they were wearing.
. The naturalness of their positions.

156

Lines 16-25

- The circumstances indicate that these people had been killed,
- because the king was regarded as divine.
- They had died calmly, accepting their fate,
- probably after taking a drug.
- They were then buried.

Undue Length

Penalties are as follows:

Summary of more than 110 words: mark out of 9
Summary of more than 120 words: mark out of 8
Summary of more than 130 words: mark out of 7
Summary of more than 140 words: mark out of 6
Summary of more than 150 words: mark out of 5
Summary of more than 160 words: mark out of 4
Summary of more than 170 words: mark out of 3
Summary of more than 180 words: mark out of 0

SCEEB/SCRE O-grade English Project
'Criterion Test'

PAPER III

'Factual' and 'Persuasive' Writing - 1 hour

INSTRUCTIONS Answer **TWO** questions, one from each section. You should spend about 30 minutes on each question.

A — 'Factual' Writing

Do **EITHER** 1 **OR** 2.

1. 'The Essentials of Jumping' might be the title of an article on show jumping. It would probably have a short introduction to catch the interest of a reader; the bulk of the article would instruct a beginner in the object of the sport and in the most important skills involved.

Choose a sport or pastime that you know well, and write a short article on it of about 200 words suitable for inclusion in a handbook of sports and pastimes.

The title might be 'The Essentials of . . . . '.

OR

2. Explain clearly and logically, as though to help a friend who didn't know how to do it, what is involved in **one** of the following:

(*a*) washing and setting or blow-drying hair.
(*b*) cooking a stew, with potatoes and vegetables of your choice.
(*c*) routine care of a dog, cat or other pet.
(*d*) playing draughts **OR** dominoes **OR** ludo **OR** snakes and ladders **OR** a common card game.

B — 'Persuasive' Writing

Do **EITHER** 1 **OR** 2 **OR** 3.

1. In the school magazine, there are to be reports on the activities of various school societies. A society you belong to needs more members; to attract them, you are asked to give the sort of report which will interest people in joining.

Write your report.

OR

2. The Headteacher of your school has asked for a statement of the senior pupils' views on the idea of a School Council, consisting of elected staff and pupils, to deal with school rules and major breaches of them.

Write your statement, arguing as persuasively as you can for or against the idea.

OR

3. Write an article for a school magazine arguing for **or** against **ONE** of the following:

   (*a*) Compulsory sports at school.
   (*b*) Cookery or Housekeeping for boys.
   (*c*) Technical Subjects for girls.
   (*d*) Reducing the school leaving age to 14.

'Factual' and 'Persuasive' Writing — Criteria

Marks 10 —————— 7½ —————— 5 —————— 2½ —————— 0

GOOD WORK	AVERAGE WORK	POOR WORK
Factual ● Thorough knowledge of subject shown. Good *selection* of most important points (especially in No. 1). All demands of the question met. *Persuasive* A definite position firmly stated. *Several* reasonable arguments in support.	*Factual* ● Knowledge of subject. Some attempt at *selection* of main points, but a tendency to try to include everything OR a tendency to omit important points. Occasional irrelevant points made. *Persuasive* Position is stated. Arguments in support fewer in number and less forceful than in good answers.	*Factual* ● Little real knowledge of the subject shown. No attempt to select main points. Omission of many important points. Much may be irrelevant.
● Good logical or chronological organisation of material. Clear paragraphing where appropriate.	● Some attempt at organisation of material, but not fully successful. Paragraphing where appropriate.	● No attempt at organisation of material – a hotchpotch of ideas/points.
● Awareness of the reader. Appropriate tone and style for the 'audience' and function defined by the question.	● Attempt to write for the specified 'audience', but not sustained, or not quite properly pitched in tone.	● No awareness shown of 'audience'.
● In '*Factual*', untrammelled expression, accurate vocabulary, clear sentence structures. In '*Persuasive*', a range of appropriate vocabulary; possibly fairly successful use of emotive language and rhetorical sentence structures.	● In '*Factual*', fairly straightforward clear vocabulary and sentence structures, perhaps with tendency to ramble or with some vagueness. In '*Persuasive*', adequate vocabulary and sentence structures; possibly attempts at emotive language and rhetorical devices.	● Failure to exhibit command of adequate vocabulary and sentence structures.
● Correctness in nearly every instance (Spelling, Punctuation and Grammar).	● Generally correct Spelling, Punctuation and Grammar; correct use of full stops almost always though with some faults.	● Many technical faults, causing the flow of communication to be impeded – e.g. complete or almost complete failure to use full stops correctly.

SCEEB/SCRE O-grade English Project
'Criterion Test'

PAPER IV

Interpretation-Multiple Choice

PASSAGE BOOKLET

INSTRUCTIONS Do not write anything on this Booklet. Read each Passage carefully. Then answer the questions by writing the **LETTER** of the correct answer next to the question number on the answer sheet.

PASSAGE 1

For absolute effrontery in selling fake antiques I have yet to come across anyone to beat Herbert Pomfret. Gentle-mannered and charming, with the slightly pathetic air of an aristocrat who has fallen on hard times and is down to his last Rolls-Royce, Herbert gives each prospective victim a talk on the perils of investing in antiques nowadays and the wicked tricks some unscrupulous villains play on people who give them their trust. It is a technique which rarely fails.

Herbert finds his victims by inserting small-ads in Personal Columns. Usually he offers a single item: a superb Louis XVI table, perhaps; a Renoir, a diamond and ruby necklace. If he is selling a painting, he quickly dismisses the victim's attempts to talk money. 'Before I could even think of discussing a price,' says Herbert, 'I want you to consult an independent expert.'

He insists on the victim removing the picture from the wall and taking it, in a hired car, to any one of the art galleries in the Bond Street area. There, while Herbert waits in the car at the parking meter, an art expert examines the picture in return for a spot consultation fee and declares the canvas to be genuine. He might say: 'I estimate its value at approximately £3,000.'

The victim returns to the waiting car with the picture and offers Herbert £2,000. Whereupon Herbert (who knows his stuff; I'll give him that), sighs, 'I was expecting the value to be nearer £3,000. Ah, well.'

He accepts the offer, and on receiving the victim's cheque makes a reasonable request. 'It will take me three days to find a suitable replacement to fill that gap on my wall,' he says. 'Also, it will allow time for your cheque to be cleared. To prove that the sale has been completed, and that the painting is yours, please be good enough to sign your name on the back of the canvas here, with my pen.'

The victim signs. Three days later he collects the canvas from a distinctly cold Mr Pomfret who points out that he has discovered the painting to be indeed worth £3,000; nevertheless, he is prepared to honour his bargain. He hands over the canvas. The victim's signature is on the back. Home goes the victim with his purchase, a little regretful, possibly, at having won a deal with a defenceless elderly gentleman, but elated at having made, he believes, £1,000 profit.

Probably not until years later, when he comes to sell his precious investment, does he discover that the painting is a forgery. Herbert had concealed it in the back of the genuine picture's frame. It was the forgery the victim signed.

At least he can't say that Herbert didn't warn him . . .

PASSAGE 2

(After many years a wife is allowed to visit her husband who is in a prison camp.)

They had only half an hour, and the seconds passed quickly, like grains of sand trickling through the neck of an hour-glass. Dozens of urgent questions crowded Natasha's mind, but all she asked was:

'When did you find out about the visit?'

'Day before yesterday. And you?'

'Tuesday.'

'How are things at work?' he enquired.

'Why do you ask?' she said anxiously. 'Do you know?'

'What do you mean?'

He knew certain things, but he wasn't sure what was in her mind.

He knew for instance, that the wives of prisoners were always being harassed. But how could he know that last Wednesday his wife had lost her job because she was married to him? Having received only three days ago a notice granting permission for this visit, she had not yet begun to look for new work. She had waited for their meeting, as though by some miracle it might give illumination, showing her what to do.

But how could he advise her? He had been in prison for too long and knew nothing of the world outside.

The great decision she had to make was whether to divorce him or not. As the minutes slipped by in this drab, poorly heated room with its dim light from the barred window, her hopes of a miracle were fading.

She realised that in this meagre half hour she could not give her husband any idea of her loneliness and suffering, and that he had a life of his own, which had taken a quite different course from hers. So why should she upset him with her worries, if they meant nothing to him anyway?

The warder turned to the wall and began to study the plaster.

'Are you sad all alone? Are you very sad?' he asked still tenderly rubbing his cheek against her hand.

'Was she sad?' he asked! Now the visit was nearly over, she felt sick at heart. She would soon have to go out into the bleak streets feeling none the better for this meeting, and utterly alone. Her dreary life enveloped her like grey cotton wool.

'Natasha!' He stroked her hands. 'If you count it up, I haven't much more to go now – only three years. Only three . . . '

'Only three!' She stopped him angrily, feeling her voice tremble, and losing control of it. 'Only three? Only, you say! I'm at the end of my tether. I won't last another month! I might as well die. The neighbours treat me like dirt – they've thrown out my trunk. I've stopped going to see my sisters and my aunt – they all jeer at me and say they've never heard of such a fool. They keep telling me to divorce you and remarry. When is all this going to end? Just look at me! I'm thirty-seven years old. In three years I'll be an old woman. I come home and I don't make myself dinner; I don't clean the room – I haven't the heart. I just flop down on the couch and lie there like a log. I beg you my darling, please do something to get out earlier.'

She had not meant to say any of this, but it was all too much for her. Shaking with sobs and kissing her husband's hand, she let her head fall against the rough warped little table, which had seen many such tears.

'Please calm yourself,' the warder said sheepishly, glancing at the open door.

The lieutenant-colonel stood grimly in the doorway, glaring at the woman's back and shut the door.

The regulations did not explicitly forbid the shedding of tears, but, if one went by the spirit of the law, they clearly could not be permitted.

PASSAGE 3

A great deal of irresponsible nonsense is written about the effects of sex, violence and bad language on television.

It is quite right that we should be concerned with the effects of these, but the fact is that detailed analytic research from Himmelweit to Halloran has failed to produce conclusive evidence to justify the imposition of external censorship over television.

BBC and ITA have detailed codes of practice and have conducted research into the effect of television violence on children. A considerable percentage of advertisements is already rejected and pressure on certain programmes is quickly effective.

The latest responsible contribution to the debate has just been made available by Southern TV, who carried out an investigation into 'good taste on television'. The aim was to assess the level of feeling among parents about sex and violence on British television, but not, it must be emphasised, to prove any cause and effect relationship.

Several hundred interviews were carried out under the direction of two trained psychologists and a battery of questions was prepared. A distinction was drawn in the conclusions between spontaneous, unprompted criticism in response to general queries and criticism emanating from direct questioning on specific points.

The conclusions do show that a considerable number of parents are concerned about the effects of violence on children, but comparatively few can recall specific, harmful incidents that have affected them or their children.

A sizable body of opinion expects harmful lasting effects on young people in general with hardliners on this in the 45-plus age-group. The main objections were fears of imitative behaviour, personal embarrassment at watching certain scenes and a feeling that television was encouraging passivity.

Probably the most important conclusion to emerge is that most parents consider the existing safeguards for the protection of children adequate. There is certainly little evidence of any general drive for the imposition of a more exacting censorship.

SCEEB/SCRE O-grade English Project
'Criterion Test'

PAPER IV

Interpretation-Multiple Choice

QUESTION BOOKLET

INSTRUCTIONS Do not write anything on this Booklet. Read each Passage carefully. Then answer the questions by writing the **LETTER** of the correct answer next to the question number on the answer sheet.

PASSAGE 1

1. Which of the following could **not** be regarded as an antique?
A An old but elegant chair
B A nineteenth century tea-pot
C A four-poster bed
D An electric clock
E A Roman coin

2. The phrase closest in meaning to 'absolute effrontery' (line 1) is
A utter deceitfulness
B thorough greed
C downright trickery
D sheer impudence
E considerable skill.

3. In the first paragraph Herbert Pomfret warns prospective buyers against unscrupulous dealers because he
A is an honest dealer himself
B wants buyers to avoid his competitors
C wants to win the buyer's confidence
D is interested in the tricks of dishonest dealers
E likes to talk business with his customers.

4. The word nearest in meaning to 'pathetic' (line 3) is
A stupid
B inadequate
C pitiful
D unhappy
E hopeless.

5. Herbert's 'technique which rarely fails' (line 7) is to
A encourage the buying of antiques
B warn against swindlers
C drive a Rolls-Royce
D advise against buying antiques
E pretend to be upper class.

6. The expression 'knows his stuff' as used in line 21 means that Herbert
A is a good judge of the value of antiques
B can recognise potential victims easily
C knows how to trick people cleverly
D is good at pretending to be knowledgeable
E can always obtain antiques when he wants them.

7. Herbert's main reason for wanting to keep the picture for three days longer is to
A obtain a similar picture to put on the wall
B copy the victim's signature on to a fake
C check on the real value of the picture
D remove the genuine picture from the frame
E ensure that the victim can actually pay.

8. The phrase '. . . a distinctly cold Mr Pomfret . . " (lines 29-30) means that Herbert
A did not care if the buyer was dishonest
B spoke in a very clear voice
C was polite but unfriendly
D was angry at being cheated
E was sorry to part with a good painting.

9. In lines 32-35 the buyer probably feels all the following **except**
A a little ashamed
B proud of his business sense
C flushed with success
D sorry for Herbert
E slightly foolish.

164

10. All the following encourage the buyer to think he is cheating the old man **except** that Herbert
   A makes him consult an independent expert
   B does not accompany him to the expert
   C accepts his valuation of £2,000
   D honours the bargain made with him
   E asks him to leave the picture for three days.

11. Which one of the following is **not** part of Herbert's method of winning the victim's confidence?
   A The fact that the art expert is in his pay
   B His willingness to have an outside evaluation
   C His cultured style and courtesy
   D His hiring a car to go to the art gallery
   E The sense of guilt he creates in the buyer

PASSAGE 2

12. The opening conversation between Natasha and her husband could best be described as
   A casual
   B angry
   C strained
   D formal
   E intimate.

13. According to lines 11-16, Natasha had hoped that her visit to her husband would
   A console her for the loss of her job
   B prevent her from being harassed
   C suggest a line of action to her
   D inform her husband that she had lost her job
   E cast light on her gloomy existence.

14. Natasha might wish to seek a divorce for all of the following reasons **except** that she
   A had been made a social outcast
   B no longer loved her husband
   C had grown tired of waiting
   D hoped to get another job
   E was desperately lonely.

15. Which one of the following best describes the attitude of Natasha's husband during the visit?
   A Loving but insensitive
   B Bitter and unfeeling
   C Inquisitive but sympathetic
   D Depressed and pessimistic
   E Anxious but hopeful

16. The husband's questions in line 27 revealed most clearly to Natasha his
   A lack of understanding of her real situation
   B sorrow at being separated from her
   C complete understanding of her unhappiness
   D unselfish nature
   E stupidity in asking obvious questions.

17. All of the following factors contribute to Natasha's unhappiness **except**
   A her sense of isolation from her husband
   B her relatives' opinion of her
   C the unkindness of those around her
   D her scorn for her husband's lack of spirit
   E her sense of the pointlessness of life.

18. The harshness of Natasha's situation is suggested in the passage by all of the following **except**
   A the use of adjectives expressing lack of colour and hope
   B her isolation from friends and relations
   C the cruelty of the warder
   D the fact that she has very little time with her husband
   E her lack of interest in her home.

19. According to the passage, which one of the following is true of the room in which the visit takes place?
   A The plaster was carved in an interesting way.
   B The electric light was switched on.
   C The furniture was rough but well made.
   D The paintwork was old and unattractive.
   E The atmosphere was hot and stuffy.

20. The warder's behaviour during the visit shows that he was
   A bashful and curious
   B ill-mannered and anxious
   C unconcerned and impatient
   D embarrassed and fearful
   E inconsiderate and bored.

21. Which one of the following adjectives best sums up what the author feels about 'the spirit of the law' (line 54)? It is
   A strict
   B unfair
   C heartless
   D unbending
   E logical.

PASSAGE 3

22. The passage suggests that much of what is written about the effects of sex, violence and bad language on television is 'irresponsible' because
- A there are definitely no harmful effects
- B it is written by people over 45
- C it is not supported by proper research findings
- D there is no need for any censorship
- E the BBC and the ITA are always watchful.

23. According to the passage, at the moment television programmes are censored by
- A the Government
- B the television organisations
- C trained psychologists
- D parents
- E nobody.

24. All these statements about the conduct of the Southern TV investigation are true **except**
- A the questions were thought out in advance
- B criticism based on specific questions was ignored
- C the researchers were interested in people's feelings about television
- D the questions were aimed only at parents
- E the investigations were carried out by means of interviews.

25. According to the passage, all of the following statements about television companies are true **except** that they
- A already have clear rules about what they can show
- B will make changes in some programmes under pressure
- C have studied the problems of children and violent programmes
- D turn away at present a sizeable proportion of advertisements
- E consider external censorship of programmes unnecessary.

26. Which one of the following is closest in meaning to 'assess' as used in line 13?
- A Keep up-to-date with
- B Take a sample of
- C Get firm control of
- D Obtain information about
- E Make less embarrassing

27. According to the passage, the stated purpose of the investigation carried out by Southern TV was to
- A find out how much sex and violence is shown on television
- B make a responsible contribution to a controversial issue
- C find out if there was a close link between television viewing and behaviour
- D discover how parents felt about the treatment of certain issues by television
- E decide whether there was a need for stricter control over television.

28. Parents interviewed made all of the following objections about TV programmes **except**
- A violent scenes might have an effect on children
- B children might copy what was shown
- C some television scenes were embarrassing
- D some favourite programmes were shown too late in the evening
- E children might be discouraged from more active occupations.

29. The Southern TV investigation indicates all of the following **except**
- A few people wish a stricter control of programmes
- B not all parents worry about the effects of television
- C most people object to any form of censorship
- D older people tend to be more critical of television
- E people's fears are not confined to the effects of violence.

30. Which two of the following sentences best summarise the main conclusions of the passage as a whole?
- (i) There is no need to impose further censorship on television.
- (ii) The amount of violence on television is on the increase.
- (iii) People over 45 years old find television more objectionable.
- (iv) Proof that television is a bad influence on children is lacking.

- A (i) and (ii) only
- B (i) and (iii) only
- C (i) and (iv) only
- D (ii) and (iii) only
- E (iii) and (iv) only

SCEEB/SCRE O-grade English Project
'Criterion Test'

PAPER V

Short Story - 1 hour 45 minutes

INSTRUCTIONS
1. The teacher will read the story aloud – follow it in the text.
2. Then read carefully page 1, headed 'Purposes of Study'.
3. Now read the story again yourself.
4. Try to answer IN YOUR HEAD the 'Leading Questions' on page 1.
5. WRITE your answers to the 'Assessment Questions' on page 2, giving as full answers as you can.
6. Please note that there is no need to hurry: you have plenty of time to think before writing your answers.

Purposes of Study

As you study **Inexperience** have the following purposes in mind:

1. *Feelings:* to find out all you can about the feelings of the girl and of Andrew, and about the reasons for them.
2. *Right or Wrong?* to judge how far *you* approve of what they say and do, and how far you think the *author* meant you to approve.
3. *The Writer's Skill:*
   (*a*) to note the writer's use of the *setting* and its relationship to the characters and the plot;
   (*b*) to note 'clues' in the story which suggest the background of the characters and their ideas and what is likely to happen as the story progresses;
   (*c*) to respond to the shifts of sympathy the author invites from you, so that you imagine yourself in the same situation as the characters at the different stages of the story.

When you have read the story again, answer the following questions in your head before going on to the written answers. Finding answers to these preliminary questions should help you to write better, more complete answers to the assessment questions.

'Leading Questions' (answers not to be written)

1. Why had Andrew put on his oldest clothes?
2. Which of the following words most accurately expresses your feeling for the girl at the beginning of the story:
   liking; contempt; sympathy; impatience?
   Justify your choice by thinking of **two** points about her.
3. Whom would you tend to blame for the fact that the two are not talking to each other in the opening paragraphs?
4. Pick out 3 or 4 hints in the story that the girl has been plotted against.

and that much of the conversation deliberately ignores her.

5. Why does Andrew explain about Tom and Maria 'carefully and unnaturally'?
6. (*a*) What becomes clearer when Andrew says, 'Well, beyond Hampstead then'?
   (*b*) What is he doing during the silence while he finishes his drink?
7. Why is the girl so slow to believe that she is being given the brush-off?
8. Why do you think the author called the story **Inexperience**?
9. What are your feelings about the girl and about Andrew at the end of the story?

Assessment Questions

Answer these questions, **as fully as you can**, in writing:

Feelings

1. Why are the girl and Andrew not speaking to each other at the beginning of the story? **(4)**
2. Trace the changes in the girl's feelings for Andrew step by step through the story. Mark each step by a phrase or sentence quoted from the story. **(6)**
3. (*a*) What arrangements have Andrew and Colin made before the story begins? **(2)**
   (*b*) Explain how successful it is and what it tells us about Andrew's feelings towards the girl. **(4)**

Right or Wrong

1. What changes have come over your attitude to the girl between the beginning and the end of the story? Explain why. **(6)**
2. What feelings towards Andrew do you have at the end of the story? Taking into account the probable reasons for what he did, explain what led you to feel as you do. **(6)**

The Writer's Skill

1. Mention three things that make us feel the girl is being plotted against. **(3)**
2. Read these three paragraphs again:
   (*a*) Colour flooded into her cheeks. She had forgotten, and she was genuinely sorry, for she did not want to be unkind to him. Earlier, at the beginning of the whole thing, she had firmly decided not to mind about his being poor (lines 52-54).
   (*b*) The suddenness of the hurt made her gasp and lose all balance for a moment; her expression became ragged and wild-looking. She had known, perhaps long ago, that they were arriving at this point. It was too soon, though, far too soon (lines 98-101).
   (*c*) This stunned her – for a moment he thought it was because she took it as a huge lie. Then he saw that she believed him (lines 139-140).

Considering **each** paragraph separately, explain how the writer:
(i) helps us to understand more about Tisha and Andrew; **(6)**
(ii) leaves 'clues' suggesting what will happen later in the story. **(3)**

40

168

Inexperience by Frank Tuohy

The girl stood with her back to the bar, slightly in everyone's way. A pretty girl, she was wearing a dress with a flared skirt that she wore at cocktail parties. She swung her foot to and fro, and looked around her at the dull groups of raincoated men, whom she took to be commercial travellers. Her glances always ended with a little gesture of irritation, as if these people and these places were too boring and typical; then she took up her glass, pretending to drink deeply. She usually sipped a very little, sometimes nothing at all. She let the beer touch her mouth and slide back into the glass. She wanted to make it last a long time.

The young man had on a patched tweed coat and service dress trousers: he was about twenty-one and had just come out of the Army. He was looking down at the top of the bar. Neither of them spoke.

Now another young man pushed his way towards the bar, where, after exchanging a few words with the landlord, he peered round about him until he saw the couple. Their expressions and their way of standing were arranged for the general public, not for each other. He could guess that they had not been talking.

'Hullo, hullo!' the newcomer said. He was tall and had a fair handsome face; his expensive overcoat hung open.

'Oh, Colin. You know Colin, don't you?'

'Of course I know Colin.' The girl turned her face, bright with her anxiety to be liked, full on him.

'Hello there, Fred.' Colin called girls 'Fred', but he seemed embarrassed. 'Have I kept you waiting?'

'Waiting?' the girl asked, puzzled, for she had not expected to see him at all. She looked at her young man, but he did not help her.

'I'm sorry,' Colin said suddenly, rousing himself. 'What are you drinking?'

'Oh, thank you,' the girl said, 'but I'm all right with this one.' The young man was frowning at Colin across her head.

'I see.' He caught the landlord's eye, and bought drinks for himself and his friend. 'What've you been doing?'

'We went to the Curzon,' the girl said helpfully. 'It was really awfully good, wasn't it Andrew?'

'You said you liked it,' her young man answered.

'Well, you liked it too, you know you did.'

'Me?' He made an incredulous voice. 'I thought it stank.'

'Then why did you take me to it, if you thought it was so awful?' she asked crossly.

To interrupt, Colin said: 'What was this film, Fred? You haven't said what it was yet.'

She told him the name of the film. 'Have you seen it?'

'Yes.'

'Well, you liked it, surely?'

Colin looked across at Andrew, laughing. 'I must say I thought it stank, too.'

The girl made a curious trapped movement of her head. 'It was jolly good,' she said, blinking. 'I thought it was jolly good.' She examined the stitching of one of her gloves and started humming a little tune to herself; then she came back at Andrew savagely: 'It's your turn to buy a drink.'

He made a hopeless face at her. 'Tisha, you know perfectly well I haven't any more money. I told you.'

Colour flooded into her cheeks. She had forgotten, and she was genuinely sorry, for she did not want to be unkind to him. Earlier, at the beginning of the whole thing, she had firmly decided not to mind about his being poor.

'Give me a cigarette then,' she said quite softly. 'Look, there's a table free, why don't we sit down?'

But she felt their resistance as she said it, and again she had the feeling that she had said the wrong thing. They remained standing, and soon afterwards the two men began a conversation about people she didn't know. Without listening to them she smoked her cigarette in little puffs, with a slight frown.

'Tom and Maria,' Colin was saying. 'They're in a pretty bad way. I've been up there all afternoon and I did say we'd go back there. I don't think they should be left alone.'

The girl looked from one to the other, her eyes half-closed, as if trying in a sophisticated way to sum them up.

'What sort of thing was happening?'

'The usual things. They'll probably be turned out of the studio.'

'I see.'

The girl watched Andrew carefully whenever he talked. She was anxious about him. It had been Sunday afternoon and she was aware that all had not gone very well. She had dressed up, but he had appeared in his oldest clothes. He had not spoken much; it became increasingly obvious to her that he reserved his humour for state occasions. One always realised things so much more deeply after the cinema: coming out of the Curzon this afternoon into the cold streets of early summer, she knew that he no longer tried to interest her. Without meaning to, she sighed; but neither of them noticed.

Colin said rather loudly: 'Yes, I told them I'd go back and bring you.'

She looked up at him. She hadn't known they were going to meet Colin this evening, and there was something not right about his being there. Nevertheless, his presence might make the evening easier.

'We must do that, then,' Andrew said. He turned to the girl and began explaining carefully and unnaturally. 'Tom and Maria – these friends of ours – you've heard me speak of them? Well, Tom's wife been writing to him again and they're both very depressed.'

'I see.'

'Colin's been out there this afternoon – he only came back again to see us. He said we'd go up there tonight.'

'Of course, why not?' Since this fitted in with her idea of being among other people this evening, she let herself agree with what he was saying. 'Well, let's go there then.'

'But, Tisha, it'll be a great bore.' His voice was quieter.

'I don't see why, they're friends of yours, aren't they?'

'But, Tisha – '

'When do we start?'

'Not you, Tisha.'

The suddenness of the hurt made her gasp and lose all balance for a moment; her expression became ragged and wild-looking. She had known, perhaps long ago, that they were arriving at this point. It was too soon, though, far too soon.

'Well – I – really!'

'But, Tisha,' he was going on saying, 'you don't know these people.'

'I know I don't. But am I so awful that I can't be shown to them? You'd think I was a pariah* or something. Look, I want to meet your friends. I know you don't think I want to, but I do.' Suddenly it was important for her to fight, though she hated herself for it.

'Try not to get into a flap. Colin has only just told me I have to go and see these people. I don't want to go, it's a long way and it'll probably be very dull.'

---

*pariah: an outcast.

'Where is it?'

'Hampstead.'

'Hampstead's not very far.'

'Well, beyond Hampstead then. Anyway, we're not going to have a gay time, a smart time. It wouldn't be at all interesting for you.'

'How do you know?' she asked him. 'You think just because I don't belong to your lot and I still live at home, that I don't know anything at all.'

'It isn't that – '

'If you mean they won't like me, why don't you say so? Is there something I've done wrong?'

'No, really, Tisha, stop this.' He gave a groan of tiredness. 'Why do you always put everything on to yourself?'

'Because I believe that's what it's all about. I don't think you're going anywhere. I think you're just trying to get rid of me.'

'Tisha, you heard Colin tell me, didn't you? You heard him say that Tom and Maria wanted to see me?'

She gave a little hard laugh, but she could not quite turn and call the other young man a liar. 'It all sounds very suspicious to me,' she said. 'Why don't you let him go by himself?' She jerked her head rudely at Colin, who was looking at himself in the mirror-glass behind the shelves of the bar. He turned, but his glance seemed to slide away from her face.

'They asked me to go too,' Andrew said. 'I can't let them down.'

'What about me? You had a previous appointment to take me out.' An idea came to her. 'I want to go out with you. Look, if it's because you haven't any money, I'll go home and borrow some from my brother.'

There was a silence while Andrew finished his drink. She watched him furiously, twisting her gloves in her hands. He put his glass down and said: 'I didn't want to tell you this, Tisha. Tom tried to commit suicide last week.'

This stunned her – for a moment he thought it was because she took it as a huge lie. Then he saw that she believed him.

'Oh, I see.' Her voice came thin and strained. 'Well, I don't see what you can do about it. He's not going to do it again, is he? He probably only did it to show off.'

'How dare you say a thing like that?'

'No, I'm sorry – I – '

'What right have you to suggest such things – you, just a spoiled little girl who talks too much.'

She looked at him and for a moment she was horrified with herself. She was sure now that she would not see him again, but she knew that she would go on loving him for some time yet.

'We ought to go,' Colin said, looking at his watch.

'I don't know where I'm going to go,' she said pathetically, 'all dressed up like this. I've said I'll be out to dinner. Please couldn't you take me with you – I'll wait outside.'

Andrew, avoiding going near her, almost shouted: 'I tell you, Tisha, it's not possible.'

Their eyes met, wounded, angry and meaningless. They stared at each other for some time.

'Oh, very well, then. Good night.'

She went towards the door, her head held up, tottering a little on black court shoes with too-high heels. They watched her until she had gone.

Andrew broke into a sniggering mock-dance.

'Oh God! Damn, damn, damn! More drink, quickly. You know, you saved me, being there. I thought I was going to give in.'

'Let's go out somewhere and get drunk, shall we?'

'Yes – look out!'

She was standing between them, her face white and shocked.

'I've left one of my gloves behind,' she said.

(Reprinted by kind permission of A. D. Peters & Co. Ltd.)

SCRE/SCEEB 'CRITERION TEST'

PAPER V – SHORT STORY

Guidance for Markers

1. In marking the answers to **Inexperience** *ignore the marks allocated to the questions*. You should give each script an *IMPRESSION MARK* out of 20. Be prepared to use the full length of the scale of marks. (If you are used to giving marks for 'Reading' questions out of 10, it does no harm to grade answers that way first and then translate them into scores out of 20.)

2. Your assessment of the quality of answers should take the following general points into account:

1. The pupil's awareness of Tisha's feelings for Andrew and of his towards her= *'FEELINGS'*.
2. The pupil's *judgment* of both Tisha and Andrew and the degree to which the pupil has been influenced in this judgment by all the information the writer gives us about the two characters= *'RIGHT OR WRONG?'*.
3. The pupil's awareness of 'clues' in the story relating to (a) the likely outcome of events (b) the characters and social background of Tisha and Andrew= *'WRITER'S SKILL'*.

Notes:

(a) *Relevance is to the general point* rather than to specific questions.

E.g., – any accurate observations about Tisha's and Andrew's feelings for each other made in answer to any of 'FEELINGS, Questions 1, 2 and 3' should be given credit; but answers which discuss only the characters' feelings for each other under 'RIGHT OR WRONG' should be regarded as irrelevant and receive no credit.

Remember that an *overall impression of the quality of the pupil's response to the story* is what is wanted.

(b) Unlimited time was allowed for this test. It is reasonable to expect *an attempt at all three sections from any pupil obtaining a pass mark*.

(c) The list of possible points given below is not exhaustive, but is probably much longer than we can reasonably expect of O-grade pupils. Roughly speaking, one might divide answers into the following three categories:

I. GOOD ANSWERS:

FEELINGS: Show awareness of Tisha's uncertainty AND anxiety AND affection AND resentment AND (probable) bitterness at the end.

Show awareness of Andrew's desire to be rid of her AND his feeling that she is tiresome and clinging, spoiled and shallow.

RIGHT OR WRONG? Show a change in attitude towards the girl, probably from a slightly *critical* or *contemptuous* attitude (She's 'green', slow on the uptake, undignified, shallow) to sympathy because she is vulnerable, the victim of a conspiracy and is humiliated.

Show a *critical* reaction to Andrew (his behaviour *cannot* be fully approved!), but tempered by some awareness of his problem: he's inexperienced too; she's hard to throw off, yet she's not 'on his wavelength' . . .

In both cases there will be a *full* statement of reasons for the judgments made.

WRITER'S SKILL: Show awareness of 3 *different* clues suggesting conspiracy.

Show understanding of *three or four* background and character points made by the paragraphs quoted, and of two or three indications given in them by the writer of the way events are tending.

II. AVERAGE ANSWERS:

FEELINGS: Answers will make the most obvious points, but omit, e.g., Tisha's uncertainty and her later resentment, and show no awareness of *why* Andrew wants rid of her.

RIGHT OR WRONG? Answers will show some awareness of how the writer has guided the readers' reactions to the characters, but may indicate that only the most obvious clues have been noted: e.g., initial dislike leading to final sympathy for Tisha, simple criticism of Andrew, with adequate reasons for the judgment, but lacking the realisation that many aspects of their personalities and circumstances are hinted at.

WRITER'S SKILL: Show awareness of two or three 'clues' about the conspiracy, but possibly not the most significant ones.

Show understanding of at least one background or character point made by each of the three paragraphs quoted, and of at least one 'clue' about later events contained in them.

III. POOR ANSWERS:

Will show a failure throughout to come to grips with the questions set and appreciation of only the surface events of the story, lacking any but the most rudimentary awareness of its emotional and social situation and of the different personalities involved.

Inexperience: List of possible points for each question

Feelings:

1. Andrew is fed up with her.
   She wasn't at ease, didn't know what to say.
   Strained relationship.
   She is a rather shallow person.
   They have little in common.
2. Puzzled
   Uncertain
   Vaguely worried about losing him
   Wants his attention
   Some affection (love)
   Suspicious
   Hurt
   Resentful
   Angry
   Shocked
   (Probably) bitter at end.

A mere list of quotations is acceptable here if it does indicate that the pupil has noted changes in her feelings towards Andrew.

3. (a) That Colin would come to be a gooseberry.
   That Colin would spin yarn about Tom and Maria.
   To get rid of Tisha so that they would go and get drunk.

   (b) Though successful, it was more embarrassing than Andrew expected (v. 1 164) *or* It wasn't *entirely* successful – she was suspicious.
   Andrew is fed up with her – didn't want to be direct – afraid she'd overcome him if he didn't have help. Lacking courage – lacking real feeling for her.

Right or Wrong?:

1. (i) sympathetic? – or appears a bit snooty, affected, shallow?

   (ii) feel she's 'green' (film dialogue)
   'a drag' (hangs onto Andrew)
   in need of support
   slow in the uptake (re brush-off)
   undignified (has to fight
   all dressed up
   said she'd be out to dinner
   'Please take me')

   (iii) sympathetic again – she's a victim of conspiracy.
   she was vulnerable.
   she is humiliated.

2. Dislike, especially for 'sniggering mock-dance'. She was probably hard to throw off – sympathy for him in this. Dislike his dishonesty, trickery. Feel that he had more experience than she – should've known better how to end the affair. Or just 'He didn't know how to do it any other way (and needed Colin's support).'

Writer's Skill:

1. Colin's arrival is pre-arranged, but Tisha doesn't know.
   Andrew frowns over her head at Colin.
   Colin clearly 'on Andrew's side' about film.
   She 'felt their resistance' (1 57) (felt excluded).
   'Colin said rather loudly', '. . . go back . . . bring you (Andrew).' (1 79)
   Whole episode of projected 'Hampstead' visit.
   Andrew turns Tisha's natural reaction against herself (11 139-147 – the 'suicide').
   The men's merriment. (11 162-164.)

2. (i) (a) Social difference indicated.
   Also she does feel genuinely for him – yet she has to *firmly decide* not to mind about his being poor.

   (b) She refuses to be brushed off because she knows in her heart that this is the feared end of the affair. She is romantic/unrealistic and vulnerable. Andrew, however clumsily, had tried to avoid causing sudden hurt.

   (c) (1) So brutal a reality sorts ill with her romantic outlook.

   (2) She is 'green' not to see through the cock-and-bull story or the reader should realise that the point is that it *is* a lie.

(iii) Andrew is jockeying for a better position. He doesn't want to be caught out; doesn't want to be direct with Tisha; is more concerned with successfully throwing her than with the truth.

(ii) (a) Their social discrepancy is already pulling them apart and will probably continue to do so

(b) It is 'the point of no return'.

(c) Andrew is on a winning tack – he can now 'manage' Tisha or – it is this lie which leads to her humiliation.

Appendix 5

Trial Marking

Appendix 6

'Visiting Moderator' Scheme

Appendix 7

Two Examples of School Ranges, Moderators' Ranges and O-grade Ranges Compared

These Appendices are not included in this publication but are available as separate items, on request, from the SCE Examination Board.

Appendix 8

Effects of Scaling in Two Schools with 'Satisfactory' Internal Assessments

1 *Characteristics of School 24 and the Effects of Scaling*
The correlations between the Folio and the other assessments at this school are very high (Folio/Criterion Test: ·89, Folio/O-grade: ·84) but one would be wrong to assume that this means a close similarity of the results for each and every pupil. In the first place the 'tail' of poor pupils makes a strong contribution to the correlation, and if those pupils who are judged as range 13 or 14 by the school are excluded from consideration (since in other schools they would not be presented) together with those who obtain an 'F' in O-grade the correlations are substantially reduced (Folio/Criterion Test: ·76, Folio/O-grade: ·71). Secondly, a graph of the Folio assessment compared to the Criterion Test (see Figure A8·1) does not show a uniform curve, and the characteristics of this picture can be made clear by identifying the class to which each pupil belongs.

Although the school was asked to produce an all-through assessment, they have not altogether succeeded in so doing. One class in particular (Class 2) stands out as (a) not succeeding in using a comparable standard to the remainder of the school as well as (b) failing to produce a satisfactory rank order. In contrast, another class (Class 3) was assessed with remarkable accuracy – way beyond that of the O-grade – though on a standard matched to the Criterion Test rather than to the less severe O-grade. In consequence, when it was attempted to standardise the Folio Marks for the entire school a variety of effects resulted. Since Class 2 was generously marked the scaling procedure led to a lower mean mark overall and an *increase* in the severity of marking of Class 3 – a very unfair effect if it were to influence the Certificates awarded to those pupils. For Class 2 itself the marks remained too high after scaling, but the major problem resulted from the initially poor rank ordering of these pupils. Rank ordering of this class was certainly not easy, since they are concentrated in bands B and C; the result of the school having made discriminations which do not correspond to those of the Criterion Test or O-grade is that pupils of apparently similar competence obtain different marks before scaling and these differences are grossly exaggerated by the scaling procedure. The members of this class would obtain scaled marks which would make them seem more dissimilar than they really are. The assessment of the top class, Class 1, was neither poor nor especially good on rank ordering – it can be taken as representative of the majority of assessments in this respect. Class 1, however, was most successfully matched to the unscaled O-grade marks on standard. Where standard is correct but agreement on ranking is not close, no benefit is achieved by any scaling procedure. An obvious improvement in reliability results, for this class, from combining the Folio and O-grade assessments, without any statistical adjustments.

Thus this one school is a microcosm in which we can find excellent ranking on the wrong standard (Class 3), poor ranking combined with a wrong standard (Class 2), and average ranking on the correct standard (Class 1) – a very useful example. Further, from scrutiny of the separate tests, it appears that these differences in assessment among the classes are almost entirely accounted for by the Folio interpretation section. (See Chapter VII.) The folio composition marks for Class 2 are slightly high, but the main disagreement of standard is on interpretation.

Figure A8.2 shows the different effects of scaling on classes 1, 2 and 3: the 'cloud' of crosses representing class 2 has been spread out so that some pupils are further away from the 'ideal' agreement of standard than before and members of class three are even more severely dealt with than previously.

2. *Characteristics of School 22 and the Effects of Scaling*

School 22 presents a quite different picture. The correlation coefficients for this school are similar to those for school 24: Folio/Criterion Test: ·84, Folio/O-grade: ·79, but the agreement on ranking was maintained throughout the school. It can be seen from Figure A8.3 that the Folio ranking in school 22 was quite good by comparison with the Criterion Test rank order. There was more scatter when Folio marks were plotted against O-grade results (Figure A8.4); but ranking is consistent, while there is a clear severity on all pupils combined with an unwillingness to give the best pupils high marks. In these circumstances the scaling procedure was helpful in raising the mean mark and pushing the best pupils up to the top of the mark scale. (See Figure A8.5.)

It is, however, also obvious from Figure A8.5 that the stretching effect of the scaling procedure exaggerates the discrepancy when there is lack of agreement on ranking.

FIGURE A 8.1 Folio in School 24 before scaling against Criterion Test

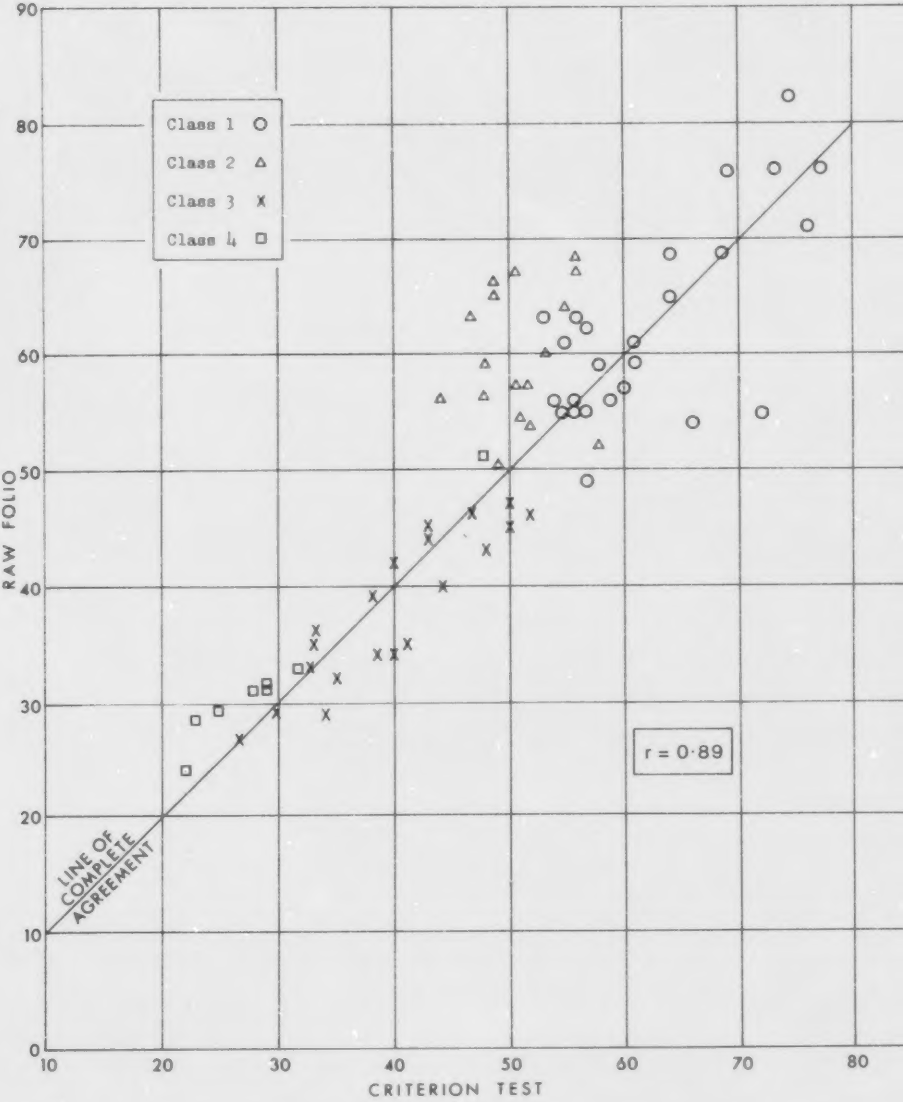
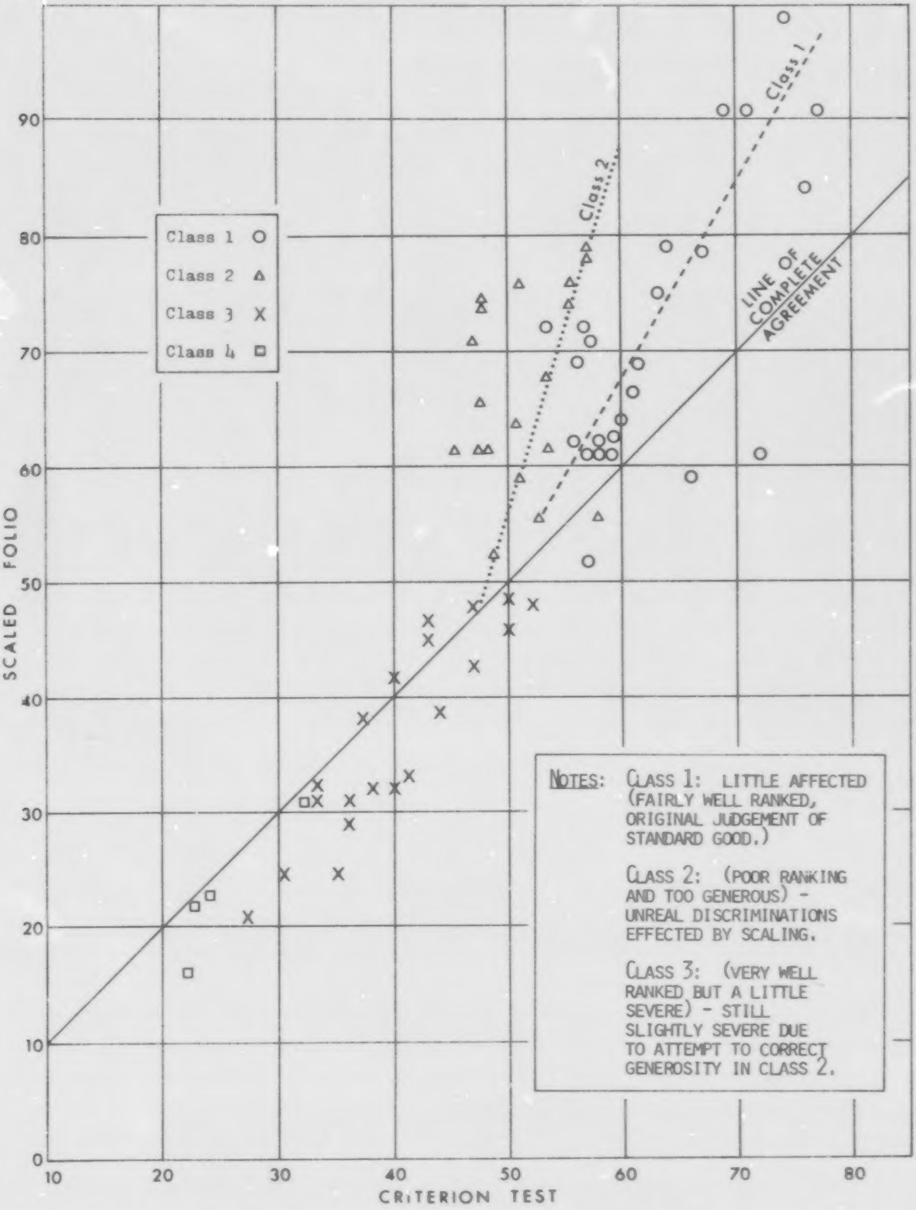


FIGURE A 8.2 Effects of scaling on classes in School 24

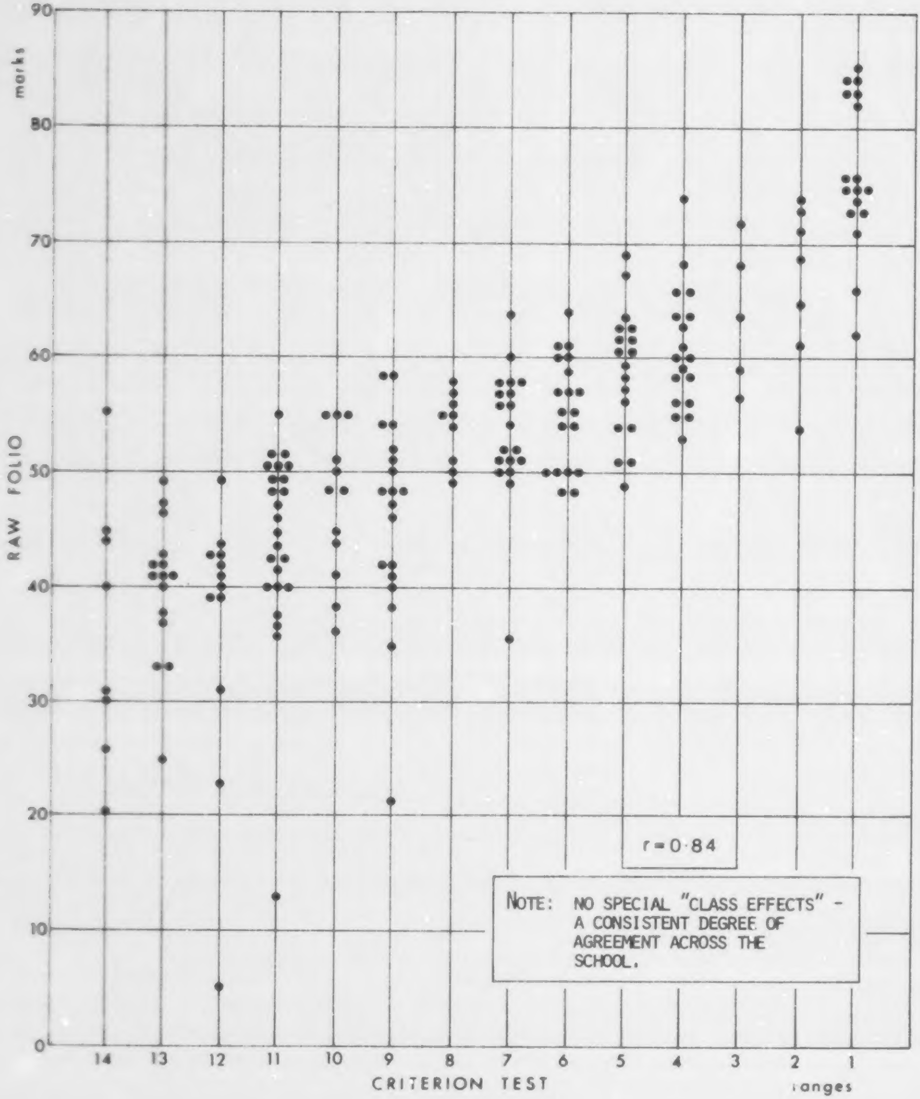


NOTES: CLASS 1: LITTLE AFFECTED (FAIRLY WELL RANKED, ORIGINAL JUDGEMENT OF STANDARD GOOD.)

CLASS 2: (POOR RANKING AND TOO GENEROUS) - UNREAL DISCRIMINATIONS EFFECTED BY SCALING.

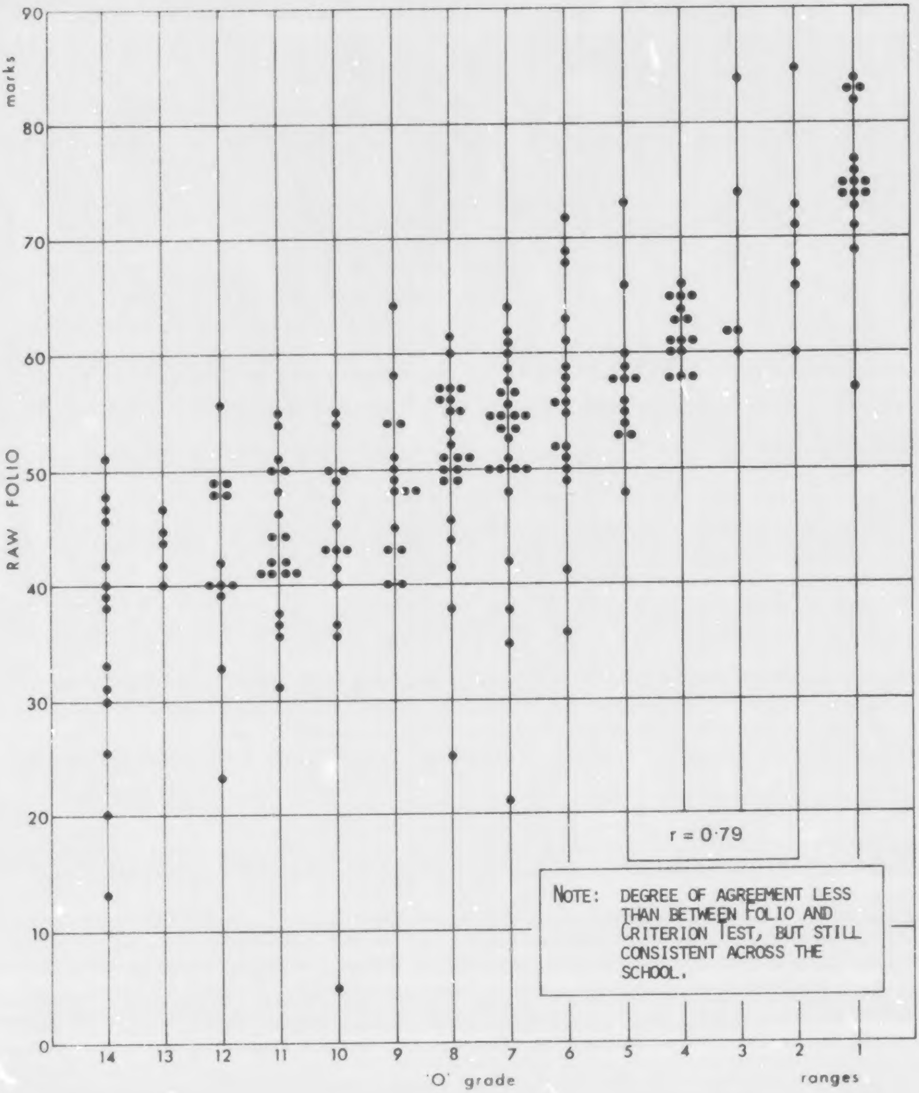
CLASS 3: (VERY WELL RANKED BUT A LITTLE SEVERE) - STILL SLIGHTLY SEVERE DUE TO ATTEMPT TO CORRECT GENEROSITY IN CLASS 2.

FIGURE A 8.3 Raw Folio against Criterion Test - School 22



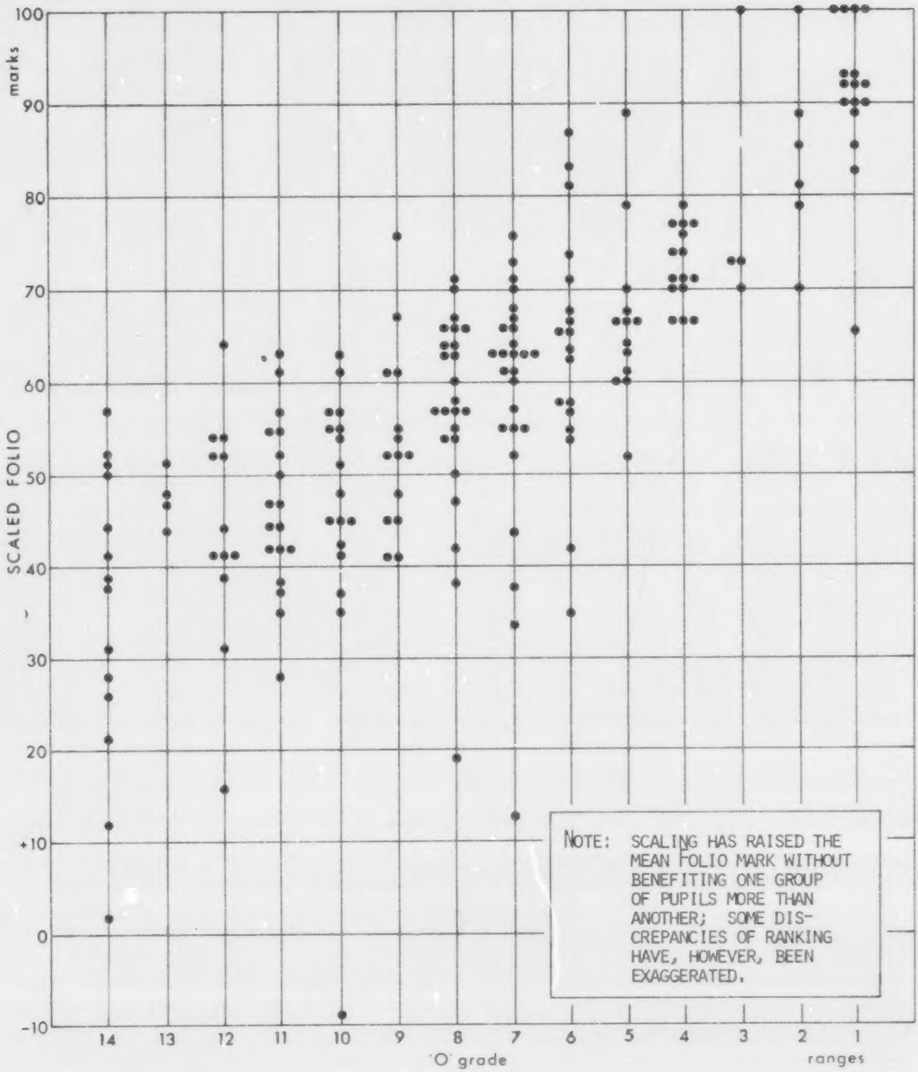
NOTE: NO SPECIAL "CLASS EFFECTS" - A CONSISTENT DEGREE OF AGREEMENT ACROSS THE SCHOOL.

FIGURE A 8.4 Raw Folio against O-grade – School 22



NOTE: DEGREE OF AGREEMENT LESS THAN BETWEEN FOLIO AND CRITERION TEST, BUT STILL CONSISTENT ACROSS THE SCHOOL.

FIGURE A 8.5 Scaled Folio against O-grade – School 22



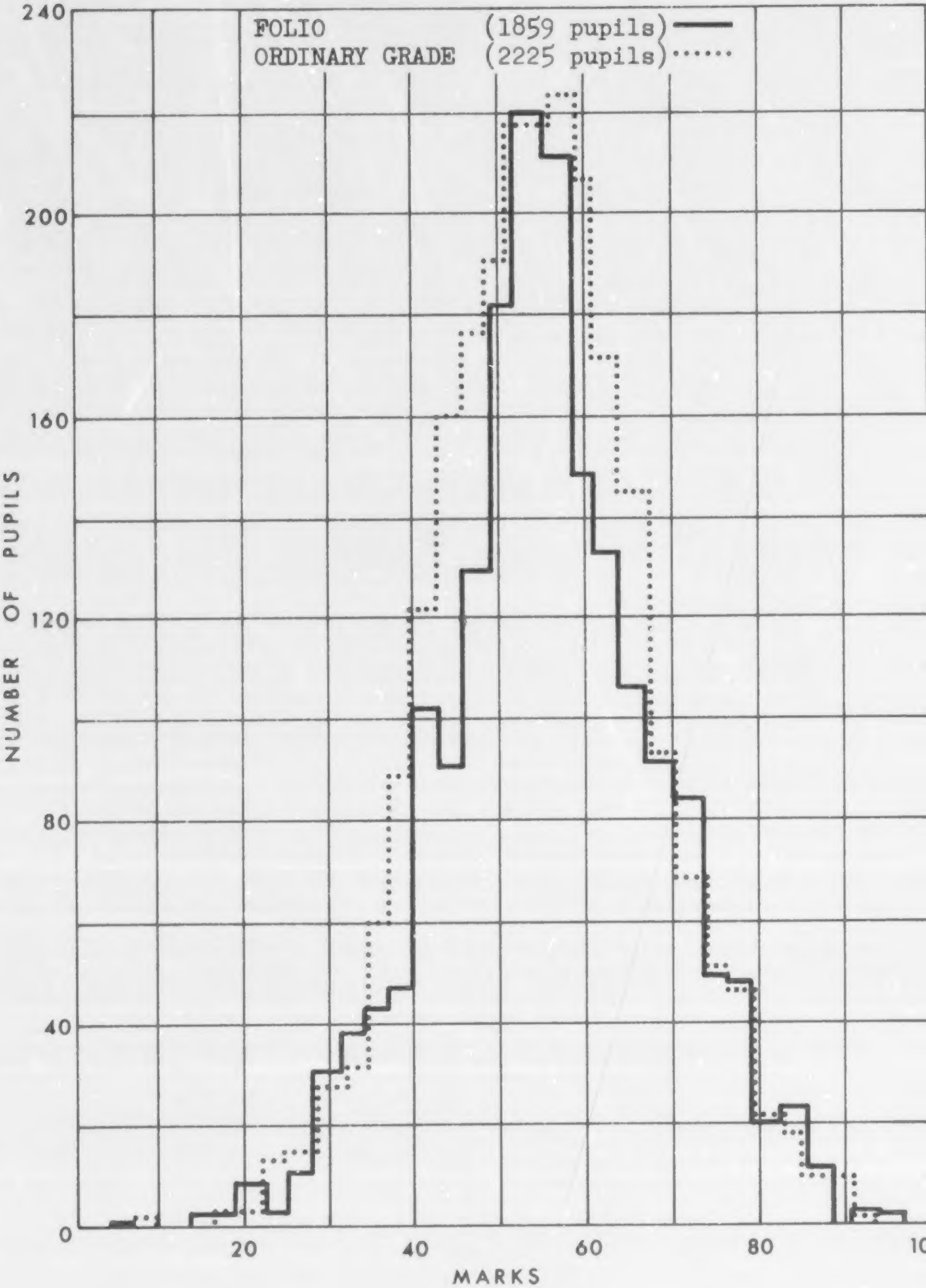
NOTE: SCALING HAS RAISED THE MEAN FOLIO MARK WITHOUT BENEFITING ONE GROUP OF PUPILS MORE THAN ANOTHER; SOME DISCREPANCIES OF RANKING HAVE, HOWEVER, BEEN EXAGGERATED.

Appendix 9

Distributions of Marks for Criterion Test Markers

This Appendix is not included in this publication but is available as a separate item, on request, from the SCE Examination Board.

Appendix 10 Distributions of Raw Marks: Folio and Ordinary grade



Appendix 11

Means and Standard Deviations (Raw Marks): O-grade, Folio and Criterion Test (possible total marks in each case = 100)

		O-grade (raw)		*Folio (raw)*		*Criterion Test (raw)*	
School	*n.*	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
11	143	57·07	9·87	57·28	10·97	51·30	8·80
13	163	53·27	12·91	55·04	11·00	45·30	11·80
14	196	54·65	12·04	54·76	9·72	47·80	11·30
15	154	53·84	11·86	55·23	14·17	49·60	11·70
21	93	57·67	9·53	61·94	10·35	48·40	11·50
22	204	57·60	12·23	52·45	12·96	50·50	12·70
23	81	54·54	13·10	53·84	14·66	45·80	13·40
24	70	55·91	12·03	52·27	13·59	50·20	12·60
32	268	53·31	11·38	55·98	11·68	47·30	11·20
33	45	54·38	9·11	55·51	8·51	50·90	9·30
35	112	59·16	11·53	57·93	10·78	53·30	12·20
All 11 schools	1529	55·36	11·78	55·51	12·01	48·90	11·80

Appendix 12
Match of Awards: Folio - O-grade (before scaling) (by school)

School / *No. of ranges mis-placed*	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9	*% within* ±1 and ±2 ranges
11 (149)		1 0·7%	0	2 1·3%	2 1·3%	11 4·4%	20 13·4%	17 11·4%	26 17·5%	29 19·5%	15 10·0%	14 9·4%	8 5·4%	3 2·0%	1 0·7%				±1=48·4% ±2=71·8%
13 (263)	1 0·4%		2 2·7%	17 6·5%	21 8·0%	25 9·5%	35 13·3%	45 17·1%	29 11·0%	25 9·5%	21 8·0%	17 6·5%	10 3·8%	5 1·9%	4 1·5%				±1=37·6% ±2=58·9%
14 (217)			2 0·9%	6 2·8%	9 4·1%	18 8·3%	28 12·9%	25 11·5%	38 17·5%	26 12·0%	32 14·7%	15 6·9%	10 4·6%	6 2·8%	2 0·9%				±1=41% ±2=68·6%
15 (163)			1 0·6%	5 3·1%	8 4·9%	16 9·8%	23 14·1%	30 18·3%	25 15·3%	19 11·6%	20 12·2%	12 7·3%	3 1·8%	1 0·6%					±1=45·2% ±2=71·4%
21 (110)				2 1·8%	4 3·6%	8 7·2%	20 18·2%	23 20·7%	24 21·6%	17 15·3%	10 9·1%	2 1·8%							±1=57·6% ±2=84·6%
22 (236)				1 0·4%	4 1·7%	9 3·8%	16 6·7%	18 7·6%	40 16·8%	40 16·8%	41 17·2%	32 13·4%	24 10·8%	6 2·5%	3 1·3%	2 0·8%			±1=41·2% ±2=65·1%
23 (109)			1 0·9%	2 1·8%	2 1·8%	5 4·6%	11 10·1%	20 18·4%	26 23·9%	18 16·5%	17 6·4%	11 10·1%	3 2·8%	2 1·8%	1 0·9%				±1=58·5% ±2=75·3%
24 (73)					1 1·4%		5 6·8%	13 17·8%	13 17·8%	14 19·2%	10 13·7%	9 12·3%	5 6·8%	2 2·7%		1 1·4%			±1=54·8% ±2=75·3%
32 (281)		2 0·7%	4 1·4%	13 4·6%	10 3·6%	29 10·3%	48 17·1%	43 15·3%	42 14·9%	30 10·7%	30 10·7%	19 6·8%	8 2·8%	3 1·1%					±1=40·9% ±2=68·7%
33 (59)				4 6·8%	3 5·1%	3 5·1%	9 15·2%	12 20·3%	12 20·3%	2 3·4%	8 13·6%	2 3·4%	3 5·1%					1 1·7%	±1=44% ±2=72·8%
35 (131)			1 0·8%	6 4·6%	4 3·1%	8 6·1%	6 4·6%	18 13·7%	14 10·7%	20 15·3%	17 13·0%	16 12·2%	10 7·6%	6 ·64%	3 2·3%	1 0·8%	1 0·8%		±1=39·7% ±2=57·3%

← Folio more lenient ‖ Folio more severe →

Appendix 13
Crosstabulation of Correlation Coefficients for all Assessment Elements in the Project (n=1529)

	Crit. Test Paper 1A	*Crit. Test Paper 1B*	*Crit. Test Paper 1*	*Crit. Test Paper 2*	*Crit. Test Paper 3*	*Crit. Test Paper 4*	*Crit. Test Paper 5*	*Crit. Test Total*	*O-grade Total*	*O-grade Paper 1*	*O-grade Paper 2*	*Folio Total*	*Folio Composition*	*Folio Interpretation*	*Folio Literature*
Crit. Test Paper 1A	1·00	0·63	0·87	0·58	0·59	0·45	0·54	0·76	0·66	0·60	0·55	0·60	0·60	0·53	0·55
Crit. Test Paper 1B	0·63	1·00	0·93	0·64	0·57	0·49	0·60	0·82	0·67	0·62	0·57	0·64	0·59	0·59	0·60
Crit. Test Paper 1	0·87	0·93	1·00	0·68	0·64	0·52	0·63	0·87	0·74	0·68	0·62	0·69	0·65	0·62	0·64
Crit. Test Paper 2	0·58	0·64	0·68	1·00	0·57	0·65	0·63	0·89	0·72	0·60	0·68	0·66	0·57	0·62	0·57
Crit. Test Paper 3	0·59	0·57	0·64	0·57	1·00	0·47	0·60	0·74	0·64	0·59	0·54	0·59	0·57	0·51	0·53
Crit. Test Paper 4	0·45	0·49	0·52	0·65	0·47	1·00	0·53	0·77	0·63	0·49	0·63	0·58	0·50	0·56	0·48
Crit. Test Paper 5	0·54	0·60	0·63	0·63	0·60	0·53	1·00	0·79	0·60	0·51	0·55	0·57	0·50	0·51	0·55
Crit. Test Total	0·76	0·82	0·87	0·89	0·74	0·78	0·79	1·00	0·82	0·70	0·75	0·76	0·68	0·70	0·67
O-grade Total	0·66	0·67	0·74	0·72	0·64	0·63	0·60	0·82	1·00	0·89	0·87	0·71	0·65	0·65	0·61
O-grade Paper 1	0·60	0·62	0·68	0·60	0·59	0·49	0·51	0·70	0·89	1·00	0·56	0·62	0·58	0·56	0·55
O-grade Paper 2	0·55	0·57	0·62	0·68	0·54	0·63	0·55	0·75	0·87	0·56	1·00	0·64	0·56	0·59	0·53
Folio Total	0·61	0·64	0·69	0·66	0·59	0·58	0·57	0·76	0·71	0·62	0·64	1·00	0·87	0·94	0·85
Folio Composition	0·60	0·59	0·65	0·57	0·57	0·50	0·51	0·68	0·65	0·58	0·56	0·87	1·00	0·70	0·73
Folio Interp.	0·53	0·59	0·62	0·62	0·51	0·56	0·51	0·70	0·65	0·56	0·59	0·94	0·70	1·00	0·70
Folio Literature	0·55	0·60	0·64	0·57	0·53	0·53	0·55	0·67	0·61	0·55	0·52	0·85	0·73	0·70	1·00

Bibliography

Adams, A. and Pearce, J. (1974) *Every English Teacher*, Oxford University Press.

Ballard, P. B. (1929) *The New Examiner* (5th impression), Hodder and Stoughton, London.

Bloomfield, B., Dobby, J. and Duckworth, D. (1977) *Mode Comparability in the CSE*, Schools Council Examinations Bulletin 36, Evans/Methuen, London.

Britton, J., Martin, N. C. and Rosen, H. (1966) *Multiple Marking of English Compositions*, Schools Council Examinations Bulletin 12, HMSO, London.

Britton, J. (1972) *What's the Use? A Schematic Account of Language Functions*, pp 245-251 of: Cashdan, A. and Grugeon, E. (editors) *Language in Education*, (Open University set book), Routledge and Kegan Paul, London.

Cohen, L. and Deale, R. N. (1977) *Assessment by Teachers in Examinations at* 16+, Evans/Methuen, London.

Hartog, P. (Chairman) (1941) *The Marking of English Essays*, Macmillan, London.

Hartog, P. and Rhodes, E. C. (1936) *The Marks of Examiners*, Macmillan, London.

Joint Matriculation Board (1965) *English Language: An Experiment in School Assessing* (*first interim report*).

Joint Matriculation Board (1967) *English Language: An Experiment in Assessing* (*second interim report*).

Joint Matriculation Board (1970) *An Experimental Scheme of School Assessment in Ordinary Level English Language: Third Report*, J.M.B., Manchester.

Lindquist, E. F. (ed) (1951) *Educational Measurement*, American Council on Education, Washington D.C.

L.A.T.E. (1965) *Assessing Compositions*, London Association for the Teaching of English, Blackie, London.

L.A.T.E. (1968), *Assessing Comprehension*, London Association for the Teaching of English, Blackie, London.

McNemar, Q. (1962) *Phsychological Statistics*, Wiley, New York.

Marcus, D. (1973) *Reports and Reporting*, The Bosworth Papers, 3 Bosworth College, Leicester.

Spencer, E. (1975) *Methods of Assessment used in making Order of Merit lists for the* 1975 *O-grade English Examination and their accuracy compared with the examination*, unpublished report to the SCEEB.

Spencer, E. (1976) *Internal Assessment or External Exam?*, Teaching English 9 No. 2, 4-9.

Stenhouse, L. (1971) *Some Limitations on the use of Objectives in Curriculum Research and Planning*, Paedogogica Europaea VI, 73-83.

Stevens, F. 1970 *English and Examinations*, Hutchison Educational, London.

Willmott, A. S. and Hall, C. G. W. (1975) *'O' Level Examined: the effect of Question Choice*, Schools Council Research Studies, Macmillan, London.

Wiseman, S. and Wrigley, J. (1958) *Essay Reliability: the effect of choice of essay title*, Educational and Psychological Measurement, 18, 129-138.

Wood, R. and Quinn, B. (1976) *Double impression marking of English Language essay and summary questions*, Educational Review 28, 229-246.

Open University (1973) *Reading Purposes, Comprehension and the Use of Context* (Reading Development, Units 3 and 4), Open University Press, Bletchley.

Open University (1977) *Developing Independence in Reading* (Reading Development, Units 5, 6, 7 and 8), Open University Press, Milton Keynes.

Robinson, F. P. (1961) *Effective Study*, Harper and Row, New York.

Schools Council (1965) *The Certificate of Secondary Education: School Based Examinations*, Schools Council Examinations Bulletin 5, HMSO, London.

SCEEB (1968) *Guidance for Teachers on the Marking of English Essays, Ordinary and Higher Grade*, Scottish Certificate of Education Examination Board, Edinburgh.

SCEEB (1974) *An Investigation into the Comparability of School Estimates and Examination Performance*, Scottish Certificate of Education Examination Board, Edinburgh.

SCRE (1976) *Pupils in Profile*, Scottish Council for Research in Education, Edinburgh, Hodder and Stoughton, London.

Scottish Education Department (1959) *Report of the Working Party on the Curriculum of the Senior Secondary School*, HMSO, Edinburgh.

Scottish Education Department (1977) *Assessment for All*, (The Dunning Report), HMSO, Edinburgh.

Smith, L. E. W. (1972) *Towards a New English Curriculum*, Dent, London.

© *Copyright S.C.E. Examination Board 1979*

Printed by
The Allen Lithographic Co. Ltd., Kirkcaldy
for the
Scottish Certificate of Education Examination Board

To be purchased from
Robert Gibson & Sons, Glasgow, Ltd.,
17 Fitzroy Place, Glasgow, G3 7SF,
or the Scottish Council for Research in Education,
16 Moray Place, Edinburgh, EH3 6DR,
or through any bookseller.